



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 113237**

**TO: Robert Kelly**  
**Location: rem-2c55**  
**Art Unit: 1632**  
**February 5, 2004**

*RCB*

**Case Serial Number: 09971773**

**From: P. Sheppard**  
**Location: Remsen Building**  
**Phone: (571) 272-2529**

**sheppard@uspto.gov**

### **Search Notes**

**THIS PAGE BLANK (USPTO)**

STIC-Biotech/ChemLib

113237

From: Kelly, Robert  
Sent: Friday, January 30, 2004 6:03 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/971,773

With regard to Application No. 09/971,773, please run the following sequence searches:

Search for SEQ ID NOS: 1, 2, 23 and 24, with results of 80% or greater sequence homology

Also, search these same sequences allowing for at least 1 insertion, deletion, substitution, or addition

Thanks!

Robert M. Kelly, Ph.D.  
Room 2C55, Remsen Bldg.  
(571) 272-0729

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 10:19:28 ; Search time 7369.83 Seconds  
(without alignments)  
11146.322 Million cell updates/sec

Title: US-09-971-773-1  
Perfect score: 2008  
Sequence: 1 aacagaaactattttctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vl.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.on.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vl.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sv.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2008	100.0	2008	6	BD168516	BD168516 Cells pro
2	1809.4	90.1	2976	10	BC010666	BC010666 Mus muscu
3	1804.6	89.9	2176	10	AB025198	AB025198 Mus muscu
4	1584	78.9	1728	6	BD168517	BD168517 Cells pro
5	1575.2	78.4	2796	9	HSAS39535	AJ539535 Homo sapi
6	1575.2	78.4	2898	9	HSY17976	Y17976 Homo sapien
7	1575.2	78.4	2902	9	HSAS39536	AJ539536 Homo sapi
8	1575.2	78.4	2992	9	HSY17977	Y17977 Homo sapien
9	1575.2	78.4	3186	9	HSY17978	Y17978 Homo sapien
10	1575.2	78.4	3280	9	HSY17979	Y17979 Homo sapien
11	1575.2	78.4	3568	9	HSAS36055	AJ536055 Homo sapi
12	1575.2	78.4	3666	9	HSAS36053	AJ536053 Homo sapi
13	1575.2	78.4	3772	9	HSAS36054	AJ536054 Homo sapi
14	1575.2	78.4	4196	9	HSAS36056	AJ536056 Homo sapi
15	1572.8	78.3	2100	6	E15725	E15725 Human mRNA
16	1571.8	78.3	2002	9	D89289	D89289 Homo sapien
17	1568	78.1	2100	6	AR170077	AR170077 Sequence
18	1514.8	75.4	1759	6	AX662735	AX662735 Sequence
19	1514.8	75.4	1759	6	BD087745	BD087745 Plant cel
20	1509.6	75.2	1836	4	D86723	D86723 Porcine mRN
21	1474.2	73.4	1728	6	AR170074	AR170074 Sequence
22	1474.2	73.4	1728	6	E14720	E14720 Procline mRN
23	1469.4	73.2	1728	4	AF247186	AF247186 Bos tauru
24	1256.4	62.6	2682	9	BC025385	BC025385 Homo sapi
25	1090.4	54.3	2129	5	XLA514872	AJ514872 Xenopus l
26	983	49.0	1850	9	AF052088	AF052088 Homo sapi
27	976.4	48.6	979	6	BD168521	BD168521 Cells pro
28	893.2	44.5	979	6	BD168522	BD168522 Cells pro
29	881	43.9	1017	6	E43925	E43925 Antibody re
30	809.8	40.3	1772	9	AB049740	AB049740 Homo sapi
31	599.8	29.9	699	6	E43926	E43926 Antibody re
C 32	446.4	22.2	110000	2	AC129041_1	Continuation (2 of
C 33	446.4	22.2	110000	2	AC129041_2	Continuation (3 of
C 34	446.4	22.2	150709	2	AC134206	AC134206 Rattus no
C 35	446.4	22.2	237954	2	AC128532	AC128532 Rattus no
C 36	441.8	22.0	844	11	BV033231	BV033231 S212P6677
C 37	441.8	22.0	171073	2	AC132342	AC132342 Mus muscu
C 38	420.4	20.9	1262	9	HSAS14325	AJ514325 Homo sapi
C 39	420.4	20.9	1396	9	HSAS14324	AJ514324 Homo sapi
40	367.6	18.3	1860	3	AF441264	AF441264 Drosophil
41	367.6	18.3	2764	3	AY051451	AY051451 Drosophil
42	336.8	16.8	2860	3	AK113018	AK113018 Ciona int
43	305.6	15.2	394	6	BD071198	BD071198 Secreted
44	302	15.0	9196	6	BD168518	BD168518 Cells pro
45	300.4	15.0	1147	9	AB049828	AB049828 Homo sapi

ALIGNMENTS

RESULT 1  
BD168516  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BD168516  
Cells producing antibody composition.  
BD168516  
WO 0231140-A/1.  
Cricetulus griseus (Chinese hamster)  
Cricetulus griseus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
1 (bases 1 to 2008)  
Kanda, Y., Sato, M., Nakamura, K., Uchida, K., Shinkawa, T., Yamane, N.,  
Hosaka, E., Yamasaki, M. and Hanai, N.

2008 bp DNA linear PAT 17-JAN-2003

TITLE Cells producing antibody composition  
JOURNAL Patent: WO 0231140-A 1 18-APR-2002;  
KYOWA HAKKO KOGYO CO LTD  
COMMENT OS Cricetulus griseus (chinese hamster)  
PN WO 0231140-A/1  
PD 18-APR-2002 WO 2001JP008804  
PF 05-OCT-2001 WO 2001JP008804  
PR 06-OCT-2000 JP OOP 308526  
PI YUTAKA KANDA, MITSUO SATO, KAZUYASU NAKAMURA, KAZUHIISA UCHIDA, PI  
TOYOHIDE SHINKAWA, NAOKO YAMANE, EMI HOSAKA, MOTOO YAMASAKI, NOBUO  
PI HANAI  
PC C12N3/10, C12P21/08, C07K16/00, A01K67/00, A61K39/395, C12N9/00, PC  
C12N15/52,  
PC G01N33/53  
CC Cells producing antibody composition  
FH Key Location/Qualifiers  
FT source 1..2008 /organism='Cricetulus griseus (chinese hamster)',  
FT Location/Qualifiers  
1..2008 /organism='Cricetulus griseus'  
/no\_type='genomic DNA'  
/db\_xref='taxon:10029'  
BASE COUNT 626 a 425 c 465 g 492 t  
ORIGIN  
Query Match 100.0%; Score 2008; DB 6; Length 2008;  
Besc Local Similarity 100.0%; Pred. No. 0;  
Matches 2008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACGAAACTTATTTTCTGTGTGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATTTCTT 60  
DB 1 AACGAAACTTATTTTCTGTGTGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATTTCTT 60  
QY 61 TGAGCTCCGAGAAGCAGAGAGGAGTGAACCTGAAATCGGAGGATGAGCTGTTCC 120  
DB 61 TGAGCTCCGAGAAGCAGAGAGGAGTGAACCTGAAATCGGAGGATGAGCTGTTCC 120  
QY 121 TGGCGTTGATATGCTCATCTTTTTCCTGGGGACCTTATCTGTTTATATAGTGGT 180  
DB 121 TGGCGTTGATATGCTCATCTTTTTCCTGGGGACCTTATCTGTTTATATAGTGGT 180  
QY 181 CATTTGGTTTCGAGATAATGACCACTGACCACTTACGAGAGAACTCTCCAAGATTTCTT 240  
DB 181 CATTTGGTTTCGAGATAATGACCACTGACCACTTACGAGAGAACTCTCCAAGATTTCTT 240  
QY 241 GCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTGAGAGAAATGGCTGAGTCTCTC 300  
DB 241 GCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTGAGAGAAATGGCTGAGTCTCTC 300  
QY 301 CGAATACAGAGGCGCTTATGATCAGGGGACAGCTACAGGAGAGTCCGTGTTTAAAG 360  
DB 301 CGAATACAGAGGCGCTTATGATCAGGGGACAGCTACAGGAGAGTCCGTGTTTAAAG 360  
QY 361 GAAAGCTGTTTAAAGCCAAAGACAGATTGAAAATTAAGAAATTAAGAAATTAAGAAATTAAG 420  
DB 361 GAAAGCTGTTTAAAGCCAAAGACAGATTGAAAATTAAGAAATTAAGAAATTAAGAAATTAAG 420  
QY 421 CTGGGAAAGGATCATGAATCTTAAAGAGGAGGATGAAAATGAGAGCTTAAAGAGCTCTGG 480  
DB 421 CTGGGAAAGGATCATGAATCTTAAAGAGGAGGATGAAAATGAGAGCTTAAAGAGCTCTGG 480  
QY 481 TTTTTCCTACAAAGTGAATTGAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 540  
DB 481 TTTTTCCTACAAAGTGAATTGAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 540  
QY 541 CATGAGATGAATCTTTTGGATTTAGGACATCATGAAGGCTTATCATGACAGATCTA 600  
DB 541 CATGAGATGAATCTTTTGGATTTAGGACATCATGAAGGCTTATCATGACAGATCTA 600  
QY 601 TACTACCTCAGTCAACAGATGGAGCTGAGTGGCGGAAATTAAGAGCTGAGTGGCGGAAATTAAG 660  
DB 601 TACTACCTCAGTCAACAGATGGAGCTGAGTGGCGGAAATTAAGAGCTGAGTGGCGGAAATTAAG 660  
QY 661 ACAGAGCTGGTCCAGCGGAGAAATTAACATATCTGAGAAATCCCAAGGACTGAGCAAAAGCC 720  
DB 661 ACAGAGCTGGTCCAGCGGAGAAATTAACATATCTGAGAAATCCCAAGGACTGAGCAAAAGCC 720  
QY 721 AGAAGCTGGTATGATTAATATCAACAGAGGCTGGCTATGGATGTCACATCTTCCATCATGTG 780  
DB 721 AGAAGCTGGTATGATTAATATCAACAGAGGCTGGCTATGGATGTCACATCTTCCATCATGTG 780  
QY 781 GTTTACTGCTTCATGATTGCTTATGGCAACCCAGCGAACACATCATCTTGGAAATCTCAGAAAT 840  
DB 781 GTTTACTGCTTCATGATTGCTTATGGCAACCCAGCGAACACATCATCTTGGAAATCTCAGAAAT 840  
QY 841 TGGCGCTATGCTACTTGGAGGATGGAGACTGTGTTTAGACCTCTGTAAGTGAGACATGACACA 900  
DB 841 TGGCGCTATGCTACTTGGAGGATGGAGACTGTGTTTAGACCTCTGTAAGTGAGACATGACACA 900  
QY 901 GACAGGTCTGGCTCTCCACTGGACACTGGTCAAGTGAAGGACAAATGTTTCAA 960  
DB 901 GACAGGTCTGGCTCTCCACTGGACACTGGTGAAGTGAAGGACAAATGTTTCAA 960  
QY 961 GTGGTCGAGCTCCCACTTTAGACAGCTCCATCTCTGCTCTCTTACTTACCTTGGCT 1020  
DB 961 GTGGTCGAGCTCCCACTTTAGACAGCTCCATCTCTGCTCTCTTACTTACCTTGGCT 1020  
QY 1021 GTACCAAGAAGCTTGCAGATCGACTCTGAGAGTCCATGTGATCTCTGAGTGGTGG 1080  
DB 1021 GTACCAAGAAGCTTGCAGATCGACTCTGAGAGTCCATGTGATCTCTGAGTGGTGG 1080  
QY 1081 GTATCCAGTTTGTCAATACTTGCCTCCCAACCTTGGCTGGAAAGGAAATAGAA 1140  
DB 1081 GTATCCAGTTTGTCAATACTTGCCTCCCAACCTTGGCTGGAAAGGAAATAGAA 1140  
QY 1141 GAAACCAACCAAGAGCTTGGCTTCAACATCCAGTATTTGAGTCCATGTGAGCGACT 1200  
DB 1141 GAAACCAACCAAGAGCTTGGCTTCAACATCCAGTATTTGAGTCCATGTGAGCGACT 1200  
QY 1201 GACAAAGTGGGAAACAGAGCGCTTCCATCCCATTTGAGGAAATACATGGTACACCGTTGAA 1260  
DB 1201 GACAAAGTGGGAAACAGAGCGCTTCCATCCCATTTGAGGAAATACATGGTACACCGTTGAA 1260  
QY 1261 GAAATTTTTCAGCTTCTGAAACGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1320  
DB 1261 GAAATTTTTCAGCTTCTGAAACGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1320  
QY 1321 ACTGATGACCTCTTCTTTTAAAGGAGGCAAGAGTACTCCAAATTAATGAATTTAT 1380  
DB 1321 ACTGATGACCTCTTCTTTTAAAGGAGGCAAGAGTACTCCAAATTAATGAATTTAT 1380  
QY 1381 AGTGATAAATCTTATTTTCTGCTGAGTGGACTACACACCGATACACAGAAATTCACCTT 1440  
DB 1381 AGTGATAAATCTTATTTTCTGCTGAGTGGACTACACACCGATACACAGAAATTCACCTT 1440  
QY 1441 CGGGGCGTGATCTGGATATACATTTCTCCAGGCTGATCTTCTTGTGTGTTT 1500  
DB 1441 CGGGGCGTGATCTGGATATACATTTCTCCAGGCTGATCTTCTTGTGTGTTT 1500  
QY 1501 TCATCCAGGCTGTAGGGTTGCTTATGAATCATGCAACACCTGATCTCTGATGCTCT 1560  
DB 1501 TCATCCAGGCTGTAGGGTTGCTTATGAATCATGCAACACCTGATCTCTGATGCTCT 1560  
QY 1561 GCAAACTTCCATCTTTTAGATGACATCTATTTTGGAGGCCAAATGCCCCAACACAG 1620  
DB 1561 GCAAACTTCCATCTTTTAGATGACATCTATTTTGGAGGCCAAATGCCCCAACACAG 1620  
QY 1621 ATTGAGTTTATCTCACCACCTCGAACTAAAGGAAATCCCATCGAACTCGAGAT 1680  
DB 1621 ATTGAGTTTATCTCACCACCTCGAACTAAAGGAAATCCCATCGAACTCGAGAT 1680  
QY 1681 ATCATTTGCTGGCTGGAAACCATTTGGTACTCTTAAAGGTTGTCAACAGAAATCTA 1740  
DB 1681 ATCATTTGCTGGCTGGAAACCATTTGGTACTCTTAAAGGTTGTCAACAGAAATCTA 1740

Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 16 Row: 1 Column: 5  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 8393370.  
 Location/Qualifiers  
 1. 2976  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="MGC:11418 IMAGE:3594582"  
 /tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating  
 ductal carcinoma, 5 month old virgin mouse."  
 /clone\_lib="NCI CGAP\_Mam6"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6"  
 1. 2976  
 /gene="Put8"  
 /db\_xref="locusID:53618"  
 /db\_xref="MGI:1858901"  
 584..2311  
 /codon\_start=1  
 /product="Fucosyltransferase 8"  
 /protein\_id="AAH10666.1"  
 /db\_xref="GI:14715013"  
 /db\_xref="locusID:53618"  
 /translation="MRAWTGSWRMILFAMGTLPIYIGHLVRDNDHPHSHRELS  
 KILAKRLQNEDELRMAESLRIPEDQGTGRVLEQLVKAKEQINRYK  
 QANGLGKHILRRIRTEAGAKLWFFLQSELKHLKLEGNELORADEILDLGHHE  
 RSTMDLYLSQTDAGDREKAKDLTELVRITVYLPQKDCSKARKLVNINKGC  
 GYCOLHVVYCFMAYGTORTLILESQNRVATGWTVPVSECTDHSGLSTGK  
 WSEVNDKN(VVELPVIDSLHPRPYLPVLAPEDLADLLRVRHGDPAVWVSQFVKY  
 LIIRPQWLEKEIBEATKLGKFKHPIGVHVRRTDKVTEAFPIETVMVHSEFQL  
 LARPMQKGVYLAATDDPTLLKEATKYSNYEISDSISWSAGLHNRVTEHSLRG  
 ILTHLSQADFLVCTFSQVAVYIMOTLHPDASANFHSLLDIYVFGQNAHQI  
 AVYHPKRTSEIPEMPGDDIIGVAGNHWGDKGNKRLKGLTKTGLYPSYKVKREKIVK  
 YPTYPEAK"

BASE COUNT 909 a 626 c 660 g 781 t

ORIGIN

Query Match 90.1%; Score 1809.4; DB 10; Length 2976;  
 Best Local Similarity 94.1%; Pred. No. 0;  
 Matches 1891; Conservative 0; Mismatches 116; Indels 2; Gaps 1;

Oy 1 AACAGAACTATTTCCTGTGCGTAACTAGAACCCAGAGTACAATGTTTCCAAATCTT 60  
 Db 483 AAAAGAAACCTATTTCTGTGTGGTAACTAGAACCCAGAGTACAATGTTTCCAGTCTT 542

Oy 61 TGAGCTCCGAGACACAGA--AGGAGCTTGAACCTCTGAAATCGGGCATGGCTGTT 118  
 Db 543 TGAGCTCCGAGAGATAGAGACAGAGTTGAAACTCTGAAATCGGGCATGGCTGTT 602

Oy 119 CCTGGCGTTGATTATGTCATTCTTTTCTGGGGGACCTTATTTTATATAGGTG 178  
 Db 603 CTGGCGTTGATTATGTCATTCTTTTCTGGGGGACCTTATTTTATATAGGTG 662

Oy 179 GTCAATTTGGTTCGAGATAATGACCACCTGACCACTTAGCAGAGAACTCTCCAGATTC 238  
 Db 663 GTCAATTTGGTTCGAGATAATGACCACCTGACCACTTAGCAGAGAACTCTCCAGATTC 722

Oy 239 TTGCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTCGAGAGAACTGCTGAGTCTC 298  
 Db 723 TTGCAAGCTTGAACGCTTAAACAAACAAATGAAGACTTCGAGAGAACTGCTGAGTCTC 782

Oy 299 TCCGAATACCAGAGGCGCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTGTTTAG 358

FEATURES  
 source  
 1. 2976  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="MGC:11418 IMAGE:3594582"  
 /tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating  
 ductal carcinoma, 5 month old virgin mouse."  
 /clone\_lib="NCI CGAP\_Mam6"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6"  
 1. 2976  
 /gene="Put8"  
 /db\_xref="locusID:53618"  
 /db\_xref="MGI:1858901"  
 584..2311  
 /codon\_start=1  
 /product="Fucosyltransferase 8"  
 /protein\_id="AAH10666.1"  
 /db\_xref="GI:14715013"  
 /db\_xref="locusID:53618"  
 /translation="MRAWTGSWRMILFAMGTLPIYIGHLVRDNDHPHSHRELS  
 KILAKRLQNEDELRMAESLRIPEDQGTGRVLEQLVKAKEQINRYK  
 QANGLGKHILRRIRTEAGAKLWFFLQSELKHLKLEGNELORADEILDLGHHE  
 RSTMDLYLSQTDAGDREKAKDLTELVRITVYLPQKDCSKARKLVNINKGC  
 GYCOLHVVYCFMAYGTORTLILESQNRVATGWTVPVSECTDHSGLSTGK  
 WSEVNDKN(VVELPVIDSLHPRPYLPVLAPEDLADLLRVRHGDPAVWVSQFVKY  
 LIIRPQWLEKEIBEATKLGKFKHPIGVHVRRTDKVTEAFPIETVMVHSEFQL  
 LARPMQKGVYLAATDDPTLLKEATKYSNYEISDSISWSAGLHNRVTEHSLRG  
 ILTHLSQADFLVCTFSQVAVYIMOTLHPDASANFHSLLDIYVFGQNAHQI  
 AVYHPKRTSEIPEMPGDDIIGVAGNHWGDKGNKRLKGLTKTGLYPSYKVKREKIVK  
 YPTYPEAK"

BASE COUNT 909 a 626 c 660 g 781 t

ORIGIN

Query Match 90.1%; Score 1809.4; DB 10; Length 2976;  
 Best Local Similarity 94.1%; Pred. No. 0;  
 Matches 1891; Conservative 0; Mismatches 116; Indels 2; Gaps 1;

Oy 1 AACAGAACTATTTCCTGTGCGTAACTAGAACCCAGAGTACAATGTTTCCAAATCTT 60  
 Db 483 AAAAGAAACCTATTTCTGTGTGGTAACTAGAACCCAGAGTACAATGTTTCCAGTCTT 542

Oy 61 TGAGCTCCGAGACACAGA--AGGAGCTTGAACCTCTGAAATCGGGCATGGCTGTT 118  
 Db 543 TGAGCTCCGAGAGATAGAGACAGAGTTGAAACTCTGAAATCGGGCATGGCTGTT 602

Oy 119 CCTGGCGTTGATTATGTCATTCTTTTCTGGGGGACCTTATTTTATATAGGTG 178  
 Db 603 CTGGCGTTGATTATGTCATTCTTTTCTGGGGGACCTTATTTTATATAGGTG 662

Oy 179 GTCAATTTGGTTCGAGATAATGACCACCTGACCACTTAGCAGAGAACTCTCCAGATTC 238  
 Db 663 GTCAATTTGGTTCGAGATAATGACCACCTGACCACTTAGCAGAGAACTCTCCAGATTC 722

Oy 239 TTGCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTCGAGAGAACTGCTGAGTCTC 298  
 Db 723 TTGCAAGCTTGAACGCTTAAACAAACAAATGAAGACTTCGAGAGAACTGCTGAGTCTC 782

Oy 299 TCCGAATACCAGAGGCGCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTGTTTAG 358

RESULT 2  
 BC010666 2976 bp mRNA linear ROD 16-APR-2003  
 LOCUS Mus musculus fucosyltransferase 8, mRNA (cdna clone MGC:11418  
 IMAGE:3594582), complete cds.  
 ACCESSION BC010666  
 VERSION BC010666.1 GI:14715012  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2976)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.W., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 2 (bases 1 to 2976)  
 Strausberg, R.  
 Direct Submission  
 Submitted (10-JUL-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center

783	Db	 TCCGAATACCAAGAGGGCCCATTTGACCAGGGGACAGCTACAGGAAGAGTCCGTGTTTTAG	842
359	Qy	 AAGAACAGCTTGTTTAAGGCCAAAGAACAGATTTGAATAATTACAAGAAACAAGCTAGGAATG	418
843	Db	 AAGAACAGCTTGTTTAAGGCCAAAGAACAGATTTGAATAATTACAAGAAACAAGCTAGGAATG	902
419	Qy	 ATCTGGGAAGAGATCATGAATCTTAAGGAGGAGGATTTGAAAATGGAGCTATAAGAGCTCT	478
903	Db	 GTCTGGGAGAGATCATGAATCTTAAGAGGAGGATTTGAATAATTGAATAATTGAAGAGCTCT	962
479	Qy	 GGTTTTTTCTACAAAGTGAAATGAAGAAATTAAGAAATTAAGAGGAAACGAACCTCCAAA	538
963	Db	 GGTTTTTTCTACAAAGCGAACTGAAGAAATTAAGAGCTTTAAGAGGAAATGAATCCAAA	1022
539	Qy	 GACATGCAGATGAATTTCTTTTGATTTTAGGACATCATGAAGGTCTATCATGACAGATC	598
1023	Db	 GACATGCAGATGAATTTCTTTTGATTTTAGGACATCATGAAGGTCTATCATGACAGATC	1082
599	Qy	 TATACTACCTCAGTCAAAACAGATCGAGTCAGTGGCGGGAAAAAGAGCCAAAGATC	658
1083	Db	 TATACTACCTCAGTCAAAACAGATCGAGTCAGTGGCGGGAAAAAGAGCCAAAGATC	1142
659	Qy	 TGACAGAGCTGGTCCAGCGGAGAAATAATATCTGCAGAAATCCCAAGGATGTCAGCAAG	718
1143	Db	 TGACAGAGCTGGTCCAGCGGAGAAATAATATCTGCAGAAATCCCAAGGATGTCAGCAAG	1202
719	Qy	 CCAGAAAGCTGGTATGTAATAATCAACAAAGCTGTGGCTATGGATGTCAACTCCATCATG	778
1203	Db	 CCAGAAAGCTGGTATGTAATAATCAACAAAGCTGTGGCTATGGATGTCAACTCCATCAG	1262
779	Qy	 TGGTTTACTGTCTCATGATTCCTTATGGCACCCAGCGAACTCATCTCTGGAAATCTCAGA	838
1263	Db	 TGGTCTACTGTTTCATGATTCCTTATGGCACCCAGCGAACTCATCTCTGGAAATCTCAGA	1322
839	Qy	 ATTGGCGCTATGCTACTCGGAGATGGGAGACTGTGTTTAGACCTGTAAAGTGAGACATGCA	898
1323	Db	 ATTGGCGCTATGCTACTCGGAGATGGGAGACTGTGTTTAGACCTGTAAAGTGAGACATGTA	1382
899	Qy	 CAGACAGGCTGGGCTCTCCACTGGACACTGGTTCAGGTGAAGTGAAGGACAAAATGTTTC	958
1383	Db	 CAGACAGATCTGGCTCTCCACTGGACACTGGTTCAGGTGAAGTGAAGGACAAAATGTTTC	1442
959	Qy	 AAGTGGTCGAGCTCCCAATGTAGACAGCCTCCATCTCGTCTCTCTTACTTACCCCTGG	1018
1443	Db	 AAGTGGTCGAGCTCCCAATGTAGACAGCCTCCATCTCGTCTCTCTTACTTACCCCTGG	1502
1019	Qy	 CTGTACCAAGACCTTTCAGATCGACTCTCGTAGAGTCCATGTGTGATCTCTCGAGTGTGGT	1078
1503	Db	 CTGTTCCAAGACCTTTCAGACCGACTCTTAAGAGTCCATGTGTGACCTCTCGAGTGTGGT	1562
1079	Qy	 GGGTATCCAGTTTGTCAATATCTTGATTCGTTCACAACTTGGCTGGAAAGGGAATAG	1138
1563	Db	 GGGTGTCCCAATTTGTCAAAATCTTGATTTGTCTCCAACTTGGCTGGAAAGGGAATAG	1622
1139	Qy	 AAGAAACCAACCAAGACTTGGCTTCAAAACATCCAGTTTATGGAGTCCATGTTCAGACGCA	1198
1623	Db	 AAGAAACCAACCAAGACTTGGCTTCAAAACATCCAGTTTATGGAGTCCATGTTCAGACGCA	1682
1199	Qy	 CTGCAAGTGGGAACAGAGCAGCCTTCCATCCCATTTAGGGAATACATGGTACAGTTG	1258
1683	Db	 CAGACAAAGTGGGAACAGAGCAGCCTTCCACCCCATCGAGGAGTACATGGTACAGTTG	1742
1259	Qy	 AAGAAACATTTTCAGCTTCTCGAACGACAGATGAAGTGGATAAAAAAGAGTGTATCTGG	1318
1743	Db	 AAGAAACATTTTCAGCTTCTCGAACGACAGATGAAGTGGATAAAAAAGAGTGTATCTGG	1802
1319	Qy	 CCACTGATGACCCCTCTTTGTTTAAAGGAGGCAAGAGCAAAAGTACTCCAAATTTATGAATTTA	1378
1803	Db	 CTACTGATGATCCCTACTTTGTTTAAAGGAGGCAAGAGCAAAAGTACTCCAAATTTATGAATTTA	1862
1379	Qy	 TTAGTGATAACTCTATTTCTTTGGTCAGCTGGACTACAAACCGGATACACAGAAAAATTCAC	1438

Db	1863	TTAGTGATAACTCTATTTCTTTGGTCAGCTGGACTACAAATCGGTACACAGAAAATTCAC	1921
Qy	1439	TTCCGGGCGGTGATCCCTGGATATACATTTCTCTCCAGGCTGACTTCTTTGTGTGTACTT	1498
Db	1923	TTCCGGGCGGTGATCCCTGGATATACATTTCTCTCACAGGCTGACTTCTTAGTGTGTACTT	1982
Qy	1499	TTTCATCCCAAGGTCGTAGGGTTGCTTATGAAATCATGCAAAACACTGCATCTCTGATGCCT	1558
Db	1983	TTTCATCCCAAGGTCGTGCGGTTGCTTATGAAATCATGCAAAACCTTGCAATCTCTGATGCCT	2042
Qy	1559	CTGCAAACTTCCATCTCTTTAGATGACATCTACTATTTTGGAGGCCAAAATGCCCAACAAC	1618
Db	2043	CTGCGAACTTCCATCTCTTTGGATGACATCTACTATTTTGGAGGCCAAAATGCCCACANATC	2102
Qy	1619	AGATTGCGAGTTTATCTCTCACCAACCTCGAACTAAAGAGGAAATCCCCATGGAACTCTGGAG	1678
Db	2103	AGATTGCTGTTTATCTCTCACAAACCTCGAACTGAAGAGGAAATTTCCAAATGGAACTCTGGAG	2162
Qy	1679	ATATCATTTGGTGTGGCTGGAAAAACATTTGGAATGGTTACTCTTAAAGGTGTCAACAGAGAAAC	1738
Db	2163	ATATCATTTGGTGTGGCTGGAAAAACATTTGGGATGGTTATCTTAAAGGTGTCAACAGAGAAAC	2222
Qy	1739	TAGGAAAAACAGGCGCTGTACCTTCTCTACAAAGTCCGAGAGAAAGATAGAAAACAGTCAAAT	1798
Db	2223	TTGGAAAAACAGGCTTATATCTCTCTCAAAAGTCCGAGAGAGAGATAGAAAACAGTCAAAGT	2282
Qy	1799	ACCTACATATCTCTGAAGCTGAAAAATAGAGATGGAGTGTAAAGAGATTAAACAAGAAATT	1858
Db	2283	ATCCCACATATCTGTAAGCTGAAAAATAGAGATGAAGTGAAGAGAGATTAAACAAGAAAT	2342
Qy	1859	TAGTTTCAGACCATCTAGCCCAACGAGAGACCCAGACTACATATGTTTTCATTTGACAGAC	1918
Db	2343	CACTTTCAGACCATCTCGGGCCAGGAGAGACGACAGACTAACACGTGTTTCTTTTGTATGAC	2402
Qy	1919	ATGCTCCGCAACCAAGACAAAGTGGGAACCTCTCAGATGCTGCATCTGGTGGAAACGCCTCTTT	1978
Db	2403	ACGCTCCACCAACCAAGACCAAGCGGGAACCTCTCAGATGCTGCATCTGGTGGAAACGCCTCTTT	2462
Qy	1979	GTGAAGGGCTGTGTGCGCTCTCAAGCCCAT	2007
Db	2463	ATGAAGGGCTGTGTGCGCTCTCAAGCCCAT	2491

```

RESULT 3
AB025198
LOCUS      2176 bp mRNA linear ROD 05-AUG-2000
DEFINITION Mus musculus mRNA for alpha-1,6-fucosyltransferase, complete cds.
ACCESSION AB025198
VERSION    AB025198.1 GI:4586553
KEYWORDS   alpha-1,6-fucosyltransferase.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS    Hayashi,H., Yoneda,A., Asada,M., Ikeita,M. and Imamura,T.
TITLE       Molecular cloning of mouse alpha-1,6-fucosyltransferase and
            expression of its mRNA in the developing cerebrum
JOURNAL     DNA Seq. 11 (1-2), 91-96 (2000)
MEDLINE     20358720
PUBMED      10902914
REFERENCE   2 (bases 1 to 2176)
AUTHORS     Hayashi,H., Yoneda,A. and Imamura,T.
TITLE       Direct Submission
JOURNAL     Submitted (19-WAR-1999) Toru Imamura, National Institute of
            Bioscience and Human Technology, Biosignaling Department; 1-1
            Higashi, Tsukuba, Ibaraki 305-8566, Japan
            (E-mail:imamura@nih.go.jp, Tel:81-298-54-6072, Fax:81-298-54-6149)
FEATURES
            Location/Qualifiers
                1..2176
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"

```

/sex="male"		/tissue type="cerebrum"		/dev stage="adult"	
205..1932		/codon_start=1		/product="alpha-1,6-fucosyltransferase"	
CDS		/protein_id="BAA76392.1"		/db_xref="GI:4586554"	
		/translation="MRAWTGSRWIMLILPAWGTLLFVIGHLVRDNDHPDHSRELS		KILAKRLKQONDELRLMAESLRIPEGPIDOGTATGRVRLVLEOLVKAKEQIENYKK	
		QARNGKDHIELRRIRENGAKELWFFLOSELKKLKHLEGNELORHADELILLDLGHE		RSIMTDYLYQFMATGADWREKADITELVORRITLYQNPDKCSKARKLVNKNKGC	
		GYGQLHVVYCFMAYTQRTILLESQNRWRYATGWEVFPVSECTDRSGLSHG		WSGVNDKNIQVLELPVDSLHPPVPLPLAPVEDLADRLLRVHGDPAVMVVSQFVK	
		LIRPQMLEKEIEBATKLGPKHPVIGVHVVRTDKVTEAFHPIEYVMVHVEOHFL		LARMOVDKKVELATDDPTLLKEANTKYSNYERISNSLSWSAGLHNRVTENSIRGV	
		ILDHFLSODPLVCTFSSQVCRVAYEIMQTLHPDASANFHSLLDDIYIFGQNAHNQI		ADYHPKPRTEEEIPMEPGDIIGVAGNHWDGYSKGINRKLKLGKTLGYPSYKVRKLETVK	
		YPTYPEAK"			
BASE COUNT		669 a	474 c	494 g	539 t
ORIGIN					
Query Match		89.9%; Score 1804.6; DB 10; Length 2176;			
Best Local Similarity		94.0%; Pred. No. 0;			
Matches 1888; Conservative		0; Mismatches 119; Indels 2; Gaps 1;			
Qy	1	AACGAAACTTATTTCTGTGTGCTAACTAGAACACAGGTACAAATGTTTCCAAATCTT	60		
Db	104	AAAGAAACCTATTTCTGTGTGCTAACTAGAACACAGGTACAAATGTTTCCAGTCTT	163		
Qy	61	TCAGCTCCGAGAAAGACAGA--AGGAGTTGAAACTCTGAAATGCGGCATGGACTGGTT	118		
Db	164	TGAGCTCAGGAGATAGAGGACAGTTGAACTCTGAAATGCGGCATGGACTGGTT	223		
Qy	119	CCTGCGTGGATATGCTCATCTTTTGTGCTGTGGGACCTTATGTTTATATATAGTG	178		
Db	224	CCTGCGTGGATATGCTCATCTTTTGTGCTGTGGGACCTTGTATTTATATATAGTG	283		
Qy	179	GTCATTTGTCGAGATATGACCACTGACCACTTCTAGCAGAGAACTCTCCAGATTC	238		
Db	284	GTCATTTGTCGAGATATGACCACTGACCACTTCTCCAGAGAACTCTCCAGATTC	343		
Qy	239	TTGCAAACTGGAGCGCTTAAACAAACAAATGAAGACTTGAGGAGAAATGGCTGAGTCTC	298		
Db	344	TTGCAAACTGGAGCGCTTAAACAAACAAATGAAGACTTGAGGAGAAATGGCTGAGTCTC	403		
Qy	299	TCCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACAGGAGAGTCCGTTTATAG	358		
Db	404	TCCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACAGGAGAGTCCGTTTATAG	463		
Qy	359	AAGACAGCTGTTAAGGCCAAGAACACAGATTGAAATTTACAGAAACAAAGCTAGGATG	418		
Db	464	AAGACAGCTGTTAAGGCCAAGAACACAGATTGAAATTTACAGAAACAAAGCTAGGATG	523		
Qy	419	ATCTGGGAAAGGATCATGAATCTTTAAGGAGGAGGATTGAAATGGAGCTTAAAGAGCTCT	478		
Db	524	GTCTGGGAAAGGATCATGAATCTTTAAGGAGGAGGATTGAAATGGAGCTTAAAGAGCTCT	583		
Qy	479	GGTTTTTCTCAAAGTGAATTAAGAAATTAAGAAATTAAGAAAGAAACGAACCTCCAAA	538		
Db	584	GGTTTTTCTCAAAGTGAATTAAGAAATTAAGAAATTAAGAAAGAAATGAACCTCCAAA	643		
Qy	539	GACATGACAGTGAATTTCTTTGGATTTAGGACATCATGAAAGGCTCTATCATGACAGATC	598		
Db	644	GACATGACAGTGAATTTCTTTGGATTTAGGACATCATGAAAGGCTCTATCATGACAGATC	703		
Qy	599	TATACCTCAGTCAAAACAGATGAGGAGGTGAGTGGCGGAAAGAAAGCAAGATC	658		
Db	704	TATACCTCAGTCAAAACAGATGAGGAGGTGAGTGGCGGAAAGAAAGCAAGATC	763		
Qy	659	TGACAGAGCTGTCCAGCGGAGAAATAACATATCTGCAGAAATCCCAGGACTGCAGCAAG	718		
Db	764	TGACAGAGCTGTCCAGCGGAGAAATAACATATCTCCAGAAATCCCAGGACTGCAGCAAG	823		
Qy	719	CCGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG	778		
Db	824	CCGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG	883		
Qy	779	TGGTTACTGCTTCATGATTTGCTTATGGCAGCCAGCAACACTCATCTTGGAAATCTCAGA	838		
Db	884	TGGTTACTGCTTCATGATTTGCTTATGGCAGCCAGCAACACTCATCTTGGAAATCTCAGA	943		
Qy	839	ATTGGCCCTATGCTTACTTGGAGGATGGGAGACTGTGTTTAGACCTGTAGTGAGACATGTA	898		
Db	944	ATTGGCCCTATGCTTACTTGGAGGATGGGAGACTGTGTTTAGACCTGTAGTGAGACATGTA	1003		
Qy	899	CAGACAGGTCTGGGCTCTCCACCTGACACACTGGTCAAGTGAAGGACAAAAATGTTC	958		
Db	1004	CAGACAGATCTGGGCTCTCCACCTGACACACTGGTCAAGTGAAGTAAATGACAAAAACATTC	1063		
Qy	959	AAGTGGTGGAGCTCCCATTTGTAGACAGCTTCATCTCTGTCCTCTTACTTACCTTTGG	1018		
Db	1064	AAGTGGTGGAGCTCCCATTTGTAGACAGCTTCATCTCTGTCCTCTTACTTACCTTTGG	1123		
Qy	1019	CTGTACCAAGAGCTTGCAGATCGACTCTCTGAGAGTCCATGTGTATCTCGAGTGTGCT	1078		
Db	1124	CTGTTCAGAGAGCTTGCAGACCGACTCTCTTAAGAGTCCATGGTGACCTCGAGTGTGCT	1183		
Qy	1079	GGGTATCCCAAGTTTGTCAAATATCTTGATCCGTCACCAACCTTGGCTGGAAAGGAAAAATAG	1138		
Db	1184	GGGTATCCCAAGTTTGTCAAATATCTTGATCCGTCACCAACCTTGGCTGGAAAGGAAAAATAG	1243		
Qy	1139	AAGAAACACCAAGAACTTGGCTTCAAAATCAGATTTATGGAGTTCATGTACAGACGCA	1198		
Db	1244	AAGAAACACCAAGAACTTGGCTTCAAAATCAGATTTATGGAGTTCATGTACAGACGCA	1303		
Qy	1199	CTGCAAGTGGGAAACAGAGCAGCTTCCATCCATTTGAGGAATACATGTACACGTTG	1258		
Db	1304	CAGCAAGTGGGAAACAGAGCAGCTTCCATCCATTTGAGGAATACATGTACACGTTG	1363		
Qy	1259	AAGAAATTTTCAAGTCTCGAAGCGAGAAATGAAGTGGATAAAAAGAGTGTATCTGG	1318		
Db	1364	AAGAAATTTTCAAGTCTCGAAGCGAGAAATGAAGTGGATAAAAAGAGTGTATCTGG	1423		
Qy	1319	CCACTGATGACCTTCTTGTAAAGAGGCAAGACAAAGTACTCCAAATATGAATTTA	1378		
Db	1424	CTACTGATGATCTTCTTGTAAAGAGGCAAGACAAAGTACTCCAAATATGAATTTA	1483		
Qy	1379	TTAGTGAATACTCTTCTTGTGTCAGTGGACTACACACCGATACACAGAAATTCAC	1438		
Db	1484	TTAGTGAATACTCTTCTTGTGTCAGTGGACTACACACCGATACACAGAAATTCAC	1543		
Qy	1439	TTGGGGCGTGATCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTACTT	1498		
Db	1544	TTGGGGCGTGATCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTACTT	1603		
Qy	1499	TTTCATCCAGGCTGTAGGGTTCCTTATGAAATCATGCAAAACCTGCATCTGATGCT	1558		
Db	1604	TTTCATCCAGGCTGTAGGGTTCCTTATGAAATCATGCAAAACCTGCATCTGATGCT	1663		
Qy	1559	CTGAAACTTCCATCTTCTTGTAGATGACATCTACTATTTTGGAGGCCAAATGCCCAACAC	1618		
Db	1664	CTGCAACTTCCATCTTCTTGTAGATGACATCTACTATTTTGGAGGCCAAATGCCCAACATC	1723		
Qy	1619	AGATTGAGTGTATCTCTTACCACCTCGAATGAAGAGGAAATCCCATGGAACTGGAG	1678		
Db	1724	AGATTGAGTGTATCTCTTACCACCTCGAATGAAGAGGAAATCCCATGGAACTGGAG	1783		
Qy	1679	ATATCATTTGGTGTGGCTTGAACCAATTTGGAAATGGTGTACTCTAAAGGTGTCAACAGAAAC	1738		
Db	1784	ATATCATTTGGTGTGGCTTGAACCAATTTGGAAATGGTGTACTCTAAAGGTGTCAACAGAAAC	1843		
Qy	1739	TAGGAAACACAGGCTGTACCTTCTTCAAAAGTCCGAGAGAGATAGAAACAGTCAAT	1798		
Db	1844	TTGAAACACAGGCTGTATCTCTTCTTCAAAAGTCCGAGAGAGATAGAAACAGTCAAT	1903		

Qy	1799	ACCTCATATCCTGAAGCTGAAATAAATAGAGATGGAGTGTAAAGAGATTAAACAACAGAATT	1855
Dd	1904	ATCCACATATCCTGAAGCTGAAAAATAGAGATGAAGTAGAACAGAGATTAAACAACAGAACT	1963
Qy	1859	TAGTTAGACCATCTCAGCCAAGCAGAGAACGCCAGACTTAACATATGGTTCAATTGCAGAC	1918
Dd	1964	CAC TTCAGACCATCTCGGCCAAGCAGAGACGCAGACTAAACAGTGCGTTTCATTTGATAGAC	2023
Qy	1919	ATGCTCCGCAACCAAGACGAAGTGCGAACCTTCAGATGCTCAGCTGGTGGAAACGCCCTCTTT	1978
Dd	2024	ACGCTCACACCAAGACGAAGCGGGAACCTTCAGATGCTCAGCTGGTGGAAACGCCCTCTTT	2083
Qy	1979	GTGAAGGGCTGCTGTGCCCCTCAAGCCCCAT	2007
Dd	2084	ATGAAGGGCTGTGCTGCCCTCAAGCCCCAT	2112
<hr/>			
RESULT 4		1728 bp DNA linear	PAT 17-JAN-2003
BD168517			
LOCUS	BD168517	Cells producing antibody composition.	
DEFINITION	BD168517		
ACCESSION	BD168517.1	GI:27874329	
VERSION	BD168517.1	WO 0231140-A/2.	
KEYWORDS		Mus musculus (house mouse)	
SOURCE		Mus musculus	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;	
REFERENCE		Kanda, Y., Sato, M., Nakamura, K., Uchida, K., Shinkawa, T., Yamane, N., Hosaka, E., Yamasaki, M. and Hanai, N. Cells producing antibody composition Patent: WO 0231140-A 2 18-APR-2002; KYOWA HAKKO KOGYO CO LTD OS Mus musculus (mouse) PN WO 0231140-A/2 PD 18-APR-2002 PF 05-OCT-2001 WO 2001JP008804 PR 06-OCT-2000 JP ODP 308526 PI YUTAKA KANDA,MITSUO SATO,KAZUYASU NAKAMURA,KAZUHISA UCHIDA, PI TOYOHIDE SHINKAWA,NAOKO YAMANAE,EMI HOSAKA,MOTOO YAMASAKI,NOBUO PI HANAII PC C12N5/10,C12P21/08,C07K16/00,A01K67/00,A61K39/395,C12N9/00, PC C12N15/52	
COMMENT		PC GOIN33/53 CC Cells producing antibody composition PH Key Location/Qualifiers FT source 1..1728 FT /organism='Mus musculus (mouse)'. Location/Qualifiers 1..1728 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090"	
FEATURES	source		
BASE COUNT	529 a	365 c	400 g
ORIGIN			
<hr/>			
Query Match	78.9%;	Score 1584;	DB 6; Length 1728;
Best Local Similarity	94.8%;	Pred. No. 0;	
Matches 1638;	Conservative	0; Mismatches 90;	Indels 0; Gaps 0;
Qy	100	ATGGGGCATGGAGCTGGTTCTCGCGTGTGATTAATGCTCAATCTTTTGGCTGGGGACC	159
Dd	1	ATGGGGCATGGAGCTGGTTCTCGCGTGTGATTAATGCTCAATCTTTTGGCTGGGGACC	60
Qy	160	TTATTGTTTTATATAGTGTGCTATTTGGTTCGAGATAATGACCACCTGACCACTTCTAGC	219
Dd	61	TTGTTATTTTATATAGTGTGCTATTTGGTTCGAGATAATGACCACCTGATCTCCAGC	120
Qy	220	AGAGAACTCTCCAGATTCTTTGCAAAGCTGGAGCGCTTAAAAACAACAAAATGAAGACTTG	279
Dd	121	AGAGAACTCTCCAGATTCTTTGCAAAGCTTGAACGCTTAAAAACAGCAAAAATGAAGACTTG	180

```
Db 1261 TACTCCATTATGAATTTATTAGTATAAATCTATTTCTTGCTCAGCTGGACTACAAAT 1320
Qy 1420 CGATACACAGAAATTCATCTCGGGCGTGTATCTGGATATACACTTCTCTCCAGGCT 1479
Db 1321 CGGTACACAGAAATTCATCTCGGGCGTGTATCTGGATATACACTTCTCTCCAGGCT 1380
Qy 1480 GACTTCTCTGTGTACTTTTTCATCCAGGCTCTGTAGGGTGTCTTATGAATCATGCAA 1539
Db 1381 GACTTCTAGTGTACTTTTTCATCCAGGCTCTGTAGGGTGTCTTATGAATCATGCAA 1440
Qy 1540 ACATGCACTCTGTACTCTGCAAACTTCATCTTTAGATGACATCTACTATTTTGA 1599
Db 1441 ACCCTGCACTCTGTACTCTGCAAACTTCATCTTTAGATGACATCTACTATTTTGA 1500
Qy 1600 GGCCAAATGCCACACACAGATTCGAGTTTATCTCTACCAACCTCGAATCTAAGAGGAA 1659
Db 1501 GGCCAAATGCCACACATCAGATTGCTTTATCTCTCAAACTCGAATCTAAGAGGAA 1560
Qy 1660 ATCCCATGGAACTGGAGATATCATTTGCTGTGGCTGGAAACCATTTGAATGGTTACTCT 1719
Db 1561 ATTCATGGAACTGGAGATATCATTTGCTGTGGCTGGAAACCATTTGGATGGTTACTCT 1620
Qy 1720 AAAGGTGCAACAGAAACTAGGAAACACAGGCTGTACCTTCTTACAAAGTCCGAGAG 1779
Db 1621 AAAGGTATCAACAGAAACTTGGAAACACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1680
Qy 1780 AAGTAGAAACAGTCAAACTACCTACATATCTCTGAAGCTGAAAAATAG 1827
Db 1681 AAGTAGAAACAGTCAAGTATCCCATATCTCTGAAGCTGAAAAATAG 1728
```

```
RESULT 5
HSAS39535 2796 bp mRNA linear PRI 03-FEB-2003
LOCUS HSAS39535 Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), clone
DEFINITION CD-L1.
ACCESSION AJ539535
VERSION AJ539535.1 GI:28200807
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1
AUTHORS Mollicone, R., Michalski, J.C., Bauvy, C., Cailliau-Thomas, A.,
Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and
Oriol, R.
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in
human embryogenesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2796)
AUTHORS Oriol, R.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Oriol R., U504, Inserm, 16 Av. Paul
Vallant-Courrier, 94807, FRANCE
FEATURES
source
1..2796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23.2"
/clone="CD-L1"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
/country="France:Paris"
1..2796
/gene="FUT8"
326..2053
/gene="FUT8"
/EC_number="2.4.1.68"
/function="adds fucose in alphanol,6 linkage to the internal
```

```
glcNAc of the core chitobiose"
/codon_start=1
/evidence=experimental
/product="alpha6-fucosyltransferase"
/protein_id="CAD62444.1"
/db_xref="GI:28200808"
/translation="MRPWTGSRWIMLILFAWGTLLFYIGHLVRDNDHPDHSRELS
KILAKLERLQONEDLRMAESLRIPGPIDQGPALGRVLEQLVKAKEQIENYKK
OTRNGLGHDEILRRRIENGAKELWFLQSELKLNLEGNLQORHADELFLDLGHKE
RSIMDLXYLSOTDAGDWEKAKOLTELQORITVYLPQKDCSKAKLVNINKGC
GYCQLHVVYCFMAYAGTORTILLESQWRATYATGWETVPRVSETCTDRSGISTGH
WSEVKDNNQVVELPVDLPRPPYPLVADLADLRVVRHDDPAVWVSVFVK
LIRPQWLEKEIEBEATKLGKFEHVHVRRTDKVTEAFHPIESIMVHVEHFOL
LARMOVKRVLATDPSLLKAEIKYNYEYISNYSWSAGLHNRVTENSLRGV
ILDIHLSQADFLVCTPSSQVCRVAYEMOTLHPDASANPHSLDDIYFGQNAHNOI
AIYAHOPRTADEIPMBPGDIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVKREKLETYK
YPTYPEAK"
polyA_signal 2766..2771
/gene="FUT8"
polyA_site 2777
BASE COUNT 874 a 552 c 605 g 765 t
ORIGIN
Query Match 78.4%; Score 1575.2; DB 9; Length 2796;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;
Qy 1 AACAGAACTTATTTCTCTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 60
Db 225 AACAGAACTTATTTCTCTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 284
Qy 61 TGAGCTCCGAGAGACACA--GAAGGAGTCTGAAACTCTGAAATCGCGGCATGGACTGGTT 118
Db 285 TGAGCTCCAGAGCTCCAGGAGAGTGGTGAATCTGAAATCTGAAATCGCGGCATGGACTGGTT 344
Qy 119 CTGCGGCTGGATTAIGCTCATTTCTTTGCTGGGGGACCTTATTTGTTTATATAGTG 178
Db 345 CTGCGGCTGGATTAIGCTCATTTCTTTGCTGGGGGACCTTCTGCTGTTTATATAGTG 404
Qy 179 GTCAATTTGTTGAGATAATGACACCCCTGACCATTTCTAGCAGAGAACTCTCCAGATTC 238
Db 405 GTCAATTTGTTGAGATAATGACACCATCTCTATCATCTAGCCGAGAACTGTCTCCAGATTC 464
Qy 239 TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTGAGGAGAAATGCTCAGTCTC 298
Db 465 TGGCAAGCTTGAAGCTTAAACCAACAGAGAAATGAAGACTTGAGGCGGAATGCCGAATCTC 524
Qy 299 TCCGAATACCAAGAGGCTCTTATTTGATCAGGGGACAGCTACAGGAAGAGTCCGTTTGTAG 358
Db 525 TCCGATACCAAGAGGCTCTTATTTGATCAGGGGACAGCTATAGGAAGAGTACGCGTTTGTAG 584
Qy 359 AAGAACAGCTTTTAAAGCCAAAGAAACAGATTGAAATTTACAGAAACAGCTAGGAATG 418
Db 585 AAGAACAGCTTTTAAAGCCAAAGAAACAGATTGAAATTTACAGAAACAGCTAGGAATG 644
Qy 419 ATCTGGAAAGGATCATGAAATCTTAAAGGAGGAGTGGAAATTTGGAGCTAAAGAGCTCT 478
Db 645 GTCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGTGGAAATTTGGAGCTAAAGAGCTCT 704
Qy 479 GGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTTAGAAAGAAACGAACTCCAAA 538
Db 705 GGTTTTTTCTACAGATGAATTTGAAGAAATTTAAAGAAATTTAGAAAGAAATGAATCTCCAA 764
Qy 539 GACATGAGATGAATTTTGTGATTTAGGACATCATGAAAGTCTATCATGACAGATC 598
Db 765 GACATGAGATGAATTTTGTGATTTAGGACATCATGAAAGTCTATGATGACGATC 824
Qy 599 TATACCTACCTCAGTCAAAACAGATGGAGGAGTGGCGGGAAGAAAGAGCCAAAGATC 658
Db 825 TATACCTACCTCAGTCAAGATGGAGGAGTGGCGGGAAGAAAGAGCCAAAGATC 884
Qy 659 TGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCAGAAATCCCAAGAGCTGCAGCAAG 718
```







KILAKUERUKQWEDLURMAESSIRIPEDGIDQGPAGRVRVLEQLVAKAEQIENYVK  
 RINGLFGJOHEILRRRANGKAEWFFLQSKLKNLGNELGRKAEFLDLDGHE  
 QSTMTLYLSTQDANGREKAEKATLQVORTLYLQPKDSKAKVLVKNRGK  
 GYQGLHHVYCPMIAYGORTLILBSONYATGVEWTFVRPSYETDSRGSITGH  
 WSEGDKNQVQVELVIDSLPHRPEPLPLAVPEADLRLVRHGDPAVMVWSQFVK  
 LIRPQWLEKEIEATKGLGFHPVIGVHVRTDKVCTEAAFPPIEYKVVHZEHQ  
 LARMQVQKRVLATDPSLKEAYKITPNYFIESONSAGHLNRITENSNGV  
 ILDHFSDQADFVLTCTQSCVRAEYIMQTLHPDPAHSFSLDIDYFQGNQAHNGI  
 AIYAHOPRTADEIPMBPGDIIIGVAGNHWDGYSKGVNRKLGRTGLYPSYKVRREKIE  
 TVYTPVBAEK

variation

```
2210 /gene="FUT8"
```

```

/gene="FUT8"

```

/note="II"

```
/frequency="
```

```
/replace="C"
```

2758. 2763

2758. : 2703  
/gene="EIT8"

```
/gene="rulo"
/note="nutat
```

2770 2775  
/noce= puca

10117  
"BULL" - "BULL"  
10117

```
/gene="F018"  
/not a="not a"
```

3959 /note="putat  
3974

2869. .2874  
10000 - "Prima"

```
/gene="FUT8"
```

```
/note="putative"
```

BASE C  
OPTIC

Query Match 78.4%: Score 1575.2: DB 9: Length 2898:

Query Match 78.4%; score 13732  
Best Local Similarity 87.5%; Pred. No. 0:

Best local similarity 0.58, seq. no. 0;  
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;

QY	1	AA	CAGAAACTTATTTTCTGTGTGCTTA	CTAGAAACCGAGTACAAATGTTTCAATCTCT	60
Db	328	AA	CAGAAGCTATATCCACTGTGCAT	TAACCTAGAAACAGAGTTACAAATGTTTCAATCTCT	387
QY	61	TG	AGCTCCGAGAGAACA--GAAGGAG	TTGAAACTCTGAAATCGCGGCATGCACTGGTT	118
Db	388	TG	AGCTCCAGGACTCCAGGGAAGT	TGAAATCTGAAATCGCGCCATGCACTGGTT	447
QY	119	CT	TGGGTTGGATTAATGCTCATCT	TTTTCCTTGGGGACCTTATCTGTTTATATAGGTG	178
Db	448	CT	TGGGTTTGAATTAATGCTCATCT	TTTTCCTTGGGGACCTTCTGCTGTTTATATAGGTG	507
QY	179	GT	CATTTCGAGATAATGACCAACC	CTGACCACTCTTAGCAGAGAACTCTCCAAGATTC	238
Db	508	GT	CATTTCGATACGAGATAATGAC	CACTCTGATCACTAGCCGAGAACTGTCCAGATTC	567
QY	239	TT	GAAAGCTGGAGCGCTTAAACA	CAAAATGAAGACTTGAGGAGAAATGGCTAGTCTC	298
Db	568	TG	CAAGCTTTAAACAGCAATGAAG	ACTTTGAGGCGAATGGCCGAATCTC	627
QY	299	T	CGAATACCAAGGCCCTATTGAT	CAGGGGACAGCTCAGGAAGAGTCCGTGTTTATG	358
Db	628	T	CGGATACCAAGGCCCTATTGAT	CAGGGGCCAGCTATAGGAAGAGTACGCGTTTATG	687
QY	359	AA	GACAGCTTTTAAAGGCCAAGAA	CAAGATTGAAATTACAAGAAACAAGCTTAGGAATG	418
Db	688	AA	GACAGCTTTTAAAGGCCAAGAA	CAAGATTGAAATTACAAGAAACAAGCTTAGGAATG	747
QY	419	AT	CTGGGAAAGGATCATGAAATCT	TTAAGGAGAGGATTGAAATTTGAGAGCTTAAGAGCTCT	478
Db	748	GT	CTGGGGAAGGATCATGAAATCT	TTAAGGAGAGGATTGAAATTTGAGAGCTTAAGAGCTCT	807
QY	479	GG	TTTTTCTCAAAAGTGAATGAAG	AAATTAAGAAATTTAAGAGGAAACGAATCTCAAA	538
Db	808	GG	TTTTTCTCAAGAGTGAATGAAG	AAATTTAAGAACTTAGAAGGAAATGAATCTCAAA	867
QY	539	GA	CTGCAGATCAAAATCTTTTTGG	ATTTAGGACATCATGAAAGTCTTATCATGACAGATC	598
Db	868	GA	CTGCAGATCAAAATCTTTTTGG	ATTTAGGACATCATGAAAGTCTTATGATGACGATC	927
QY	599	TA	TACTACTCTAGTCAAAACAGAT	GAGCGAGTGGCGGGAAAAAAGAGCCAAAAGATC	658



QY 839 ATTGGCGTATGCTACTGGAGATGGGAGACTGTGTTTATAGACCTGTGAAGTGAGACATGCA 898  
Db 1065 ATTGGCGTATGCTACTGGTGTGGAGACTGTATTATAGCCCTGTGAAGTGAGACATGCA 1124  
QY 899 CAGACAGGTCTGGCTCTCCACTGGACACTCGGTGAGGTGAAGTGAAGGACAAAATGTTTC 958  
Db 1125 CAGACAGATCTGGCATCTCCACTGGACACTGGTGGAGTGAAGTGAAGGACAAAATGTTTC 1184  
QY 959 AAGTGGTCTGAGCTCCCACTGTAGACAGCCTCCATCTCGTCCCTCTACTTACCCCTGG 1018  
Db 1185 AAGTGGTCTGAGCTCCCACTGTAGACAGCTCTCATCCCGCTCCATATTATACCCCTGG 1244  
QY 1019 CTGTACCAAGAGACCTTGCAGATCGACTCTCGAGAGTCCATGGTGCATCTCGAGTGGT 1078  
Db 1245 CTGTACCAAGAGACCTCGAGATCGACTGTACAGTGCATGGTGCATCTCGAGTGGT 1304  
QY 1079 GGGTATCCAGTTTGTCAAACTACTGTATCTCTCAACACTTGGCTGGAAGGGAATAG 1138  
Db 1305 GGGTGTCTCAGTTGTCAAACTACTGTATCTCGCCACAGCCTTGGCTAGAAAAGAAATAG 1364  
QY 1139 AAGAAACCAAGAGCTGGCTTCAACATCCAGTATTGGAGTCCATGTCAGACGCA 1198  
Db 1365 AAGAAAGCCCAAGAGCTTGGCTTCAAACTCCAGTATTGGAGTCCATGTCAGACGCA 1424  
QY 1199 CTGACAAAGTGGGAACAGACAGCCTTCCATCCATTCAGGGAATACATGTTACACGTTG 1258  
Db 1425 CAGACAAAGTGGGAACAGAGCTGGCTTCCATCCATTCAGGGAATACATGTTACATGTTG 1484  
QY 1259 AAGAACTATTTTTCAGTCTTCGAGCGCAGAAATGAAGTGGATTAAGAAAGAGTGTATCTGG 1318  
Db 1485 AAGAACTATTTTTCAGTCTTCGAGCGCAGAAATGCAAGTGGACAAAAGAGTGTATTTGG 1544  
QY 1319 CCACGTGATGACCTCTTTTGTAAAGGAGGCAAGACAAAGTACTCCAAATTTATGAATTTA 1378  
Db 1545 CCACAGATGACCTCTTTTATTAAGGAGGCAAAACAAAGTACTCCCAATTTATGAATTTA 1604  
QY 1379 TTAGTGATACTCTATTTCTTGGTTCAGTCTGACGTACACACCGCATACACAGAAATTCAC 1438  
Db 1605 TTAGTGATACTCTATTTCTTGGTTCAGTCTGACGTACACACCGCATACACAGAAATTCAC 1664  
QY 1439 TTCGGGGCGTGATCTCGGATATACACTTTCTCTCCAGGCTGACTTCTTGTGTGACTT 1498  
Db 1665 TTCGTGGAGTGATCTCGGATATACACTTTCTCTCTCAGCGAGACTTCTCTAGTGTGACTT 1724  
QY 1499 TTTGATCCAGGTCTGTAGAGTGTGTTATGAATCATGCAACACTGCTGATCTGATGCT 1558  
Db 1725 TTTGATCCAGGTCTGTAGAGTGTGTTATGAATTTATGAAATTCACAACTGATCTGATGCT 1784  
QY 1559 CTGCAAACTCCATCTTTTATGATGATCATCTACTATTTTGGAGGCCAAATGCCACACAC 1618  
Db 1785 CTGCAAACTCCATCTTTTATGATGATCATCTACTATTTTGGGGGCCAGAAATGCCACATC 1844  
QY 1619 AGATTGCAAGTTTATCTCCACCAACCTCGAACTAAAGAGAGAAATCCCAATGGAACCTGGAG 1678  
Db 1845 AAATTGCAATTTATGCTCACCACCCCGAATGCGAGATGAATTTCCCATGGAACCTGGAG 1904  
QY 1679 ATATCATTTGGTGTGCTGGAACCACTTGGATGGTACTCTTAAGGTGTCAACAGAAAC 1738  
Db 1905 ATATCATTTGGTGTGCTGGAATCATTTGGATGGTACTCTTAAGGTGTCAACAGAAAT 1964  
QY 1739 TAGGAAACACAGGCTGTACCTTCTCAAGTCCGAGAGAGATGAGAAACAGTCAAT 1798  
Db 1965 TGGGAAGCAGGCGCTATATCCCTCTCAAGTTCGAGAGAGATGAGAAACGCTCAAGT 2024  
QY 1799 ACCCTACATATCTCAAGCTGAGAAATAGAGATGAGTGTGAAGATTAATGAGTAA 1848  
Db 2025 ACCCACATATCTGAGGCTGGAATTAAGTCTGATGGAAGATTAACGACCAAACT 2084  
QY 1849 -----CAACAGAAATTTAGTTTCAGACCATCTTCAGCCCAAGCAGAACCCAG 1893  
Db 2085 CAGTTTCGACCAAACTCAGTTTCAACACCATTTTGACCAAACTGTAGATGAAGAGGGCTCTGA 2144

QY 1894 ----ACTAACATATGTTTATGACAGACATGCTCCGACCAAGAGCAAGTGGGAACCT 1949  
Db 2145 TCTAACAAAAATAAGTTATATAGTAGATAGTATCTCTCAGCACCAGAGAGCTGGGAACCTGA 2204  
QY 1950 CAGATGCTGC-ACTGTGGGAACGCTCTTTGTGAAGGCTGCTGTGCTCCCTCAAGCC 2004  
Db 2205 CATAGGCTTCAATGTGGGAATTCCTCTTTAAACAGGCTGCAATGCTCATACCC 2260  
RESULT 8  
HSY17977  
LOCUS HSY17977 2992 bp mRNA linear PRI 17-SEP-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase,  
transcript B2.  
ACCESSION Y17977  
VERSION Y17977.1 GI:3451264  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Caillaue,A., Balanzino,L., Candelier,J.J., Oriol,R. and  
Mollicon,R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2992)  
AUTHORS Caillaue,A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Caillaue, INSERM U178, 16 av P.V.  
Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
source  
1..2992  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/isolate="transcript B2"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q33"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..2992  
/gene="FUT8"  
1..101  
/gene="FUT8"  
/note="B"  
102..200  
/gene="FUT8"  
/note="C"  
201..2156  
/gene="FUT8"  
/note="D"  
429..2156  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76986.1"  
/db\_xref="GI:3451265"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPTGSRWMLILFAWGTLLFYIGHLVRDNDHPDHSRELS  
KILAKRLQONEDLRMAESLRIPGPDIDQGAIGRVVLEQLVKAKQIENYKK  
RTRNGLKDHEILRRRIENGAKLWFFLQSELKKNLKNLQRLQKHADEFLLDHHH  
RSIMDLIYLSQTDAGDWREKEADLTQLVRRITLYLQNPDCSKAKKLVNTNKG  
GYGQLHHVYCFMIVTQRTLLSQNRMYATGWTVPFVPSVETCTDRSGISTGH  
WGSVKDNQVVELPIVDSLHPRPVPLAVPDLRLVRVHGDPAVMVSPVKY  
LIRQPMLEKIEEATKLGKFPVIGVHVRTKVTGAFAHPLEEMVHVEHFOL  
LARMQDKKIVLATDPSLLEAKTKIPNTEFISNSISWAGLNRYTENSRLGV  
ILDTHFLSQAFVLCTFSSQVCRVAYEIMQTLHPDASANFSLDDIYFGGONAHQI  
AIYAHQPTATPEIPMEPGDIIGVAGNHWGYSKGNRKLGRGLYPSYKVRKEIETVK  
YPTYPEAK"

variation	2218	/gene="FUT8"	
		/note="II"	
		/frequency=".5"	
		/replace="c"	
polyA_signal	2758 - 2763	/gene="FUT8"	
		/note="putative"	
polyA_signal	2770 - 2775	/gene="FUT8"	
		/note="putative"	
polyA_signal	2869 - 2874	/gene="FUT8"	
		/note="putative"	
BASE COUNT	905 a	597 c	663 g 827 t
ORIGIN			
Query Match	78.4%;	Score 1575.2;	DB 9; Length 2992;
Best Local Similarity	87.5%;	Pred. No. 0;	
Matches 1781; Conservative	0;	Mismatches 223;	Indels 32; Gaps 4;
Qy	1	AACGAAATCTATTTCCTGTGTGCTAACTAGAAACACAGAGTACAAATGTTTCCAAATTCCT	60
Db	328	ACAGAGTCTATTACCTGTGCTACCTAAGTACAGAGTACAAATGTTTCCAAATTCCT	387
Qy	61	TGAGCTCCGAGAGACA--GAAGGAGTTGAACTCTGAAATCGGGCATGGAGTCTT	118
Db	388	TGAGCTCCAGGACTCCAGGAAAGTGAGTTGAAATCTGAAATCGGGCATGGAGTCTT	447
Qy	119	CTGGCGTTGGATTGCTCATTCTTTTGGCTGGGGACCTTATGTTTATATAGTG	178
Db	448	CTGGCGTTGGATTGCTCATTCTTTTGGCTGGGGACCTTATGTTTATATAGTG	507
Qy	179	GTCATTGTGTCAGATAATACCACTGACCATTTAGCAGAGAACTCTCCAAAGATTC	238
Db	508	GTCATTGTGTCAGATAATGACCATCTGATCATTAGCGAGAACTGTCGAATTC	567
Qy	239	TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTGAAGGAGATGCTGAGTCTC	298
Db	568	TGGCAAGCTTGAACGCTTAAACCAACAGCAGATGAAGACTTGAAGGAGATGCTGAGTCTC	627
Qy	299	TCCGATACCAAGGCCCTATTGATCAGGGACAGCTACAGGAGAGTCCGTGTTTAG	358
Db	628	TCCGATACCAAGGCCCTATTGATCAGGGACAGCTATAGGAGAGATGCGCGTTTAG	687
Qy	359	AAGAACAGCTTCTTAAGGCCAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG	418
Db	688	AAGACAGCTTCTTAAGGCCAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG	747
Qy	419	ATCTGGGAAAGGATCATGAAATCTTTAAGGAGAGGATTGAAATGGAGCTTAAAGAGCTCT	478
Db	748	GTCTGGGAAAGGATCATGAAATCTCTGAGGAGGATTGAAATGGAGCTTAAAGAGCTCT	807
Qy	479	GGTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTAAGAAACGAACTCCAAA	538
Db	808	GGTTTTTCTACAGAGTGAATTAAGAAATTAAGAAATTAAGAAACGAACTCCAAA	867
Qy	539	GACATGCAGATGAAATCTTTTGGATTAGGACATCATGAAAGGCTTATCATGACAGATC	598
Db	868	GACATGCAGATGAAATCTTTTGGATTAGGACATCATGAAAGGCTTATCATGACAGATC	927
Qy	599	TATACCTCAGTCAACAGATGAGCAGGTGAGTGGCGGAAAGAAAGCCAAAGATC	658
Db	928	TATACCTCAGTCAACAGATGAGCAGGTGAGTGGCGGAAAGAAAGCCAAAGATC	987
Qy	659	TCACAGAGTGGTCCAGCGGAGATAACATATCTCCAGAAATCCCAAGGACTGCGAGCAAG	718
Db	988	TCACAGAGTGGTCCAGCGGAGATAACATATCTCCAGAAATCCCAAGGACTGCGAGCAAG	1047
Qy	719	CCAGAAAGCTGGTATGTAATCAACAAAGGCTGTGGCTATGGCTGTGCTCCATCATG	778
Db	1048	CCAGAAAGCTGGTATGTAATCAACAAAGGCTGTGGCTATGGCTGTGCTCCATCATG	1107
Qy	779	TGGTTTACTGCTTCATGATTGCTTATGCGACCCAGCGAAACACTCATCTTGGAAATCTCAGA	838
Db	1108	TGGTTTACTGCTTCATGATTGCTTATGCGACCCAGCGAAACACTCATCTTGGAAATCTCAGA	1167
Qy	839	ATTGGCGCTATGCTACTGCGAGGATGGGAGACTGTGTTTATAGACCTGTGAAGTGAACATGCA	898
Db	1168	ATTGGCGCTATGCTACTGCGAGGATGGGAGACTGTGTTTATAGGCTGTGAAGTGAACATGCA	1227
Qy	899	CAGACAGGCTTGGCTCTCCACTGCGACACTGCGTCAAGTGAAGTGAAGCAAAATGTTTC	958
Db	1228	CAGACAGGCTTGGCTCTCCACTGCGACACTGCGTCAAGTGAAGTGAAGCAAAATGTTTC	1287
Qy	959	AAGTGTGCGAGCTCCCACTTGTAGACAGCTCCATCTCGTCCCTCTTACTTACCTCTGG	1018
Db	1288	AAGTGTGCGAGCTCCCACTTGTAGACAGCTCCATCTCCCTCCCTCATTTATCCCTTGG	1347
Qy	1019	CTGTACCAAGAGACCTTTCAGATCGACTCTCTGAGAGTCCATGCTGCTGCTGCTGCTG	1078
Db	1348	CTGTACCAAGAGACCTTTCAGATCGACTCTGTACAGTGCATGCTGCTGCTGCTGCTG	1407
Qy	1079	GGGTATCCAGTGTGCTCAAACTATGATTCGCTCCAACTCTGCTGCTGCTGCTGCTGCTG	1138
Db	1408	GGGTATCCAGTGTGCTCAAACTATGATTCGCTCCAACTCTGCTGCTGCTGCTGCTGCTG	1467
Qy	1139	AAGAAACCAACAGAGCTTGGCTTCAAACTATGATTCGCTCCAACTCTGCTGCTGCTGCTG	1198
Db	1468	AAGAAACCAACAGAGCTTGGCTTCAAACTATGATTCGCTCCAACTCTGCTGCTGCTGCTG	1527
Qy	1199	CTGACAAAGTGGGAAACAGAGCAGCTTCCATCCATTTGAGGAATACATGTCACAGTTC	1258
Db	1528	CAGACAAAGTGGGAAACAGAGCAGCTTCCATCCATTTGAGGAATACATGTCACAGTTC	1587
Qy	1259	AAGAAACATTTTCAGCTTCTCGAACGAGAAATGAAGTGGATTAAGAAAGAGTGTATCTGG	1318
Db	1588	AAGAAACATTTTCAGCTTCTCGAACGAGAAATGAAGTGGATTAAGAAAGAGTGTATCTGG	1647
Qy	1319	CCACTGATGACCTTCTTGTAAAGGAGGCAAGACAAAGTACTCCAAATTTATGAAATTTA	1378
Db	1648	CCACTGATGACCTTCTTGTAAAGGAGGCAAGACAAAGTACTCCAAATTTATGAAATTTA	1707
Qy	1379	TTAGTGATTAATCTTATTTCTTGTGCTGACCTGACATACACACCGATACACAGAAATTTAC	1438
Db	1708	TTAGTGATTAATCTTATTTCTTGTGCTGACCTGACATACACACCGATACACAGAAATTTAC	1767
Qy	1439	TTGGGCGCTGATCCTCGATATACACTTCTCTCCAGGCTGACTTCTGCTGCTGCTGCTG	1498
Db	1768	TTGGGCGCTGATCCTCGATATACACTTCTCTCTCAGGCGAGCTTCTGCTGCTGCTGCTG	1827
Qy	1499	TTTCATCCAGGCTGTAGGGTTCCTTATGAAATCATGCAAACTGCACTGCTGCTGCTGCTG	1558
Db	1828	TTTCATCCAGGCTGTAGGGTTCCTTATGAAATCATGCAAACTGCACTGCTGCTGCTGCTG	1887
Qy	1559	CTGCAAACTTCCATCTTTTATAGATGACATCTACTATTTTGGAGGCCAAATGCCCAACAC	1618
Db	1888	CTGCAAACTTCCATCTTTTATAGATGACATCTACTATTTTGGAGGCCAGAAATGCCCAACATC	1947
Qy	1619	AGATTGCAAGTTATCTCTCAACCACTCGAACTAAAGAGGAAATCCCATGAGAACTGAG	1678
Db	1948	AAATTGCAATTTATGCTCACCACCCGAACTGAGATGAAATTTCCCATGAGAACTGAG	2007
Qy	1679	ATATCATTTGCTGGCTGGAAACCATTTGGAATGTTTACTCTTAAAGGCTGCAACAGAAAC	1738
Db	2008	ATATCATTTGCTGGCTGGAAATCATTTGGATGGCTTATTTAAAGGCTGCAACAGAAAT	2067
Qy	1739	TAGGAAAAACAGGCTGTACCTTCTCAAAAGTCCGAGAGAGATAGAAACAGTCAAAAT	1798
Db	2068	TAGGAAAAACAGGCTGTATATCCCTCTCAAAAGTTCGAGAGAGATAGAAACAGTCAAGT	2127
Qy	1799	ACCTACATATCTCTGAAAGCTGAAAAATAGAGATGGAGTGTAAAGATTTAA-----	1848
Db	2128	ACCTACATATCTCTGAGGCTGAGAAATAAGAGCTCAGATGGAAGAGATTAACGCAAACT	2187
Qy	1849	-----CAACAGAAATTTAGTTTCAGACCATCTCAGCCCAAGCAGAGACCCAG-----	1893

```
Db 2188 CAGTTCGACCAACTCAGTTCAAACTTTGAGCCAAACTAGATGAAGGGCTCTGA 2247
Qy 1894 ----ACTACATATGTTTCATTCAGACATGCTCCGACCAAGAGCAAGTGGGACCCCT 1949
Db 2248 TCTAACAAAATAAGGTTATATAGTAGATATCTCTCAGCACCAAGAGCAGCTGGGAACTGA 2307
Qy 1950 CAGATGCTGC-ACTGCTGGAACGCTCTTTGTGAAGGGCTGCTGTCCTCAAGCC 2004
Db 2308 CATAGGCTTCAATTGGTGGATTCTCTTTAACAGGGCTGCAATGCTCATACCC 2363

RESULT 9
HSY17978
LOCUS HSY17978 3186 bp mRNA linear PRI 06-NOV-1998
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase
transcript Al.
ACCESSION Y17978
VERSION Y17978.1 GI:3451266
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Cailleau,A., Balanzino,L., Candelier,J.J., Oriol,R. and
Mollicone,R.
TITLE Differential splice variants of human FUT8 embryonic cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3186)
AUTHORS Cailleau,A.L.V.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V.
Couturier, 94807 Villejuif Cedex, FRANCE
COMMENT Related entry: D89289.
FEATURES
Location/Qualifiers
1..3186
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23"
/clone="FUT8 transcript Al"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
1..3186
/gene="FUT8"
1..389
/gene="FUT8"
/note="A"
390..488
/gene="FUT8"
/note="C"
489..5244
/gene="FUT8"
/note="D"
717..2444
/gene="FUT8"
/EC_number="2.4.1.68"
/note="allele 1"
/codon_start=1
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CAA76987.1"
/db_xref="GI:3451267"
/db_xref="GPTREMBL:O00235"
/translation="MRPWTGSRWIMLILFAWGTLLFYIGHLVRDNDHPDHSRELS
KILAKLERKQNEDLRMAESLRIPGPIDQGAIGRVRLVLEQLVAKEQIENYKK
QTRNLGKHDEILRPERENGAKELWFLQSELKKNLEGNELORHAEFLLDI-GHHE
RSIMIDLXLSQDAGDWREKEAKDLTELQRRITYLQNPDKCSKAKLVCNINKGC
GYGCOLHHVYCFMAYGTQRTILLESQNRWYATGWETFRPVSCTDRSGISTGH
WSGEYKQNVQVLEPIVDLSLHPPPLYPFLAVPEDLRLVRHGDPAWVWVSQFVKY
LIRPOPWEKIEEATKKLGFHPVIGVHVRRTDKVTEAAFPHEEYMWVHVBHFOL
LARRMQVDKRVYLATDPSLLKEAKTKYPNVEFISDMSISWSAGLHNRYTENSILRGV
```

```
variation
2506
/gene="FUT8"
/note="11"
/frequency=".5"
/replacement="c"
3046..3051
/gene="FUT8"
/note="putative"
3058..3083
/gene="FUT8"
/note="putative"
3157..3162
/gene="FUT8"
/note="putative"

BASE COUNT 910 a 710 c 721 g 840 t 5 others
ORIGIN
Query Match 78.4%; Score 1575.2; DB 9; Length 3186;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;

Qy 1 AACGAAACTTATTTCCTGTGTGCTTAACACTAGAACAGAGTACAATGTTTCCAAATCTT 60
Db 616 AACGAAAGTCTATTTCCTGTGCACTAACTAGAAACAGAGTTACAATGTTTCAATCTT 675
Qy 61 TGAGCTCGGAGAGACA-GAAGGAGTTGAACTCTGAAATCGGGCATGGACTGTT 118
Db 676 TGAGCTCCAGACTCCAGGGAAGTGAGTTGAAATCTGAAATCGGGCATGGACTGTT 735
Qy 119 CTTGCGGTGATTAATGCTCAATCTTTTTCCTGCGGACCTTATGTTTATATAGTG 178
Db 736 CTTGCGGTGATTAATGCTCAATCTTTTTCCTGCGGACCTTATGTTTATATAGTG 795
Qy 179 GTCAATTTGGTTCGAGATAATGACCACTGACCATCTCTGATCACTCTAGCCGAGAACTG 238
Db 796 GTCACTTTGGTACGAGATAATGACCATCTCTGATCACTCTAGCCGAGAACTG 855
Qy 239 TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTTGAGGAGAACTGCTGAGTCTC 298
Db 856 TGGCAAGCTTTGAACGTTTAAACAGCAGATGAAGACTTTGAGGCGAATGCGCGAATCTC 915
Qy 299 TCCGAATACCAAGAGGCTCTATTTCATCAGGGGACAGCTTACAGGAGAGTCCGCTGTTT 358
Db 916 TCCGATACCAAGAGGCTCTATTTCATCAGGGGACAGCTTATAGGAGAGTCCGCTGTTT 975
Qy 359 AAGAACAGCTTTGTTAAAGCCAAAGAACAGATTGAAATTTACAAGAAACAACTAGGAATG 418
Db 976 AAGAGCAGCTTTGTTAAAGCCAAAGAACAGATTGAAATTTACAAGAAACAGACCAAGATG 1035
Qy 419 ATCTGGGAAGGATCATGAAATCTTAAGGAGAGAGATTGAAATTTGAGAGCTTAAAGCTCT 478
Db 1036 GTCTGGGAGAGGATCATGAAATCTCTGAGGAGAGAGATTGAAATTTGAGAGCTTAAAGCTCT 1095
Qy 479 GGTCTTTCTTACAAAGTGAATTTGAAGAAATTAAGAAATTAAGAAACCACTCCAAA 538
Db 1096 GGTCTTTCTTACAAAGTGAATTTGAAGAAATTAAGAAATTAAGAAACCACTCCAAA 1155
Qy 539 GACATGAGATGAAATTTTCTTTGATTTAGGACATCATGAAAGGTCTATCATGACAGATC 598
Db 1156 GACATGAGATGAAATTTTCTTTGATTTAGGACATCATGAAAGGTCTATATATGAGGATC 1215
Qy 599 TATACCTACCTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGAAAGAAAGCCAAAGATC 658
Db 1216 TATACCTACCTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGAAAGAAAGCCAAAGATC 1275
Qy 659 TGACAGAGCTGGTCCAGUGGAGATAACATATCTGCAGAAATCCCAAGACATCCGACCAAG 718
Db 1276 TGACAGAACTGGTTCAGCGGAGATAACATATCTTCAAGATCCCAAGACATCCGACCAAG 1335
Qy 719 CCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCATCATG 778
```

Db 1336 CCAAAAAGCTGGTGTGTAATATCAACAAAGGCTGGCTATGGCTGTGAGCTCCATCATG 1395  
Qy 779 TGGTTTACTGCTTCATGATGCTTATGGCACCAGCAACACTCATCTTTGGAATCTCAGA 838  
Db 1396 TGGTCTACTGCTTCATGATGCTTATGGCACCAGCAACACTCATCTTTGGAATCTCAGA 1455  
Qy 839 ATTGGCGCTATGCTACTGAGAGTGGGAGACTGTGTTTAGACCTGTAAGTGAGACATGCA 898  
Db 1456 ATTGGCGCTATGCTACTGAGAGTGGGAGACTGTGTTTAGACCTGTAAGTGAGACATGCA 1515  
Qy 899 CAGACAGCTGGCCCTCTCCACTGACACACTGCTGAGTGAAGTGAAGACAAAATGTTT 958  
Db 1516 CAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGTGAAGTGAAGACAAAATGTTT 1575  
Qy 959 AAGTGGTGGAGCTCCCTCCCTGAGACAGCTCCATCTCTGCTCTCTTACTTACCTTGG 1018  
Db 1576 AAGTGGTGGAGCTCCCTCCCTGAGACAGCTCCATCTCTGCTCTCTTACTTACCTTGG 1635  
Qy 1019 CTGTACCAAGAACCTTGAGATCGACTCTCTGAGAGTCCATGGTGATCTCTGAGTGGT 1078  
Db 1636 CTGTACCAAGAACCTTGAGATCGACTCTCTGAGAGTCCATGGTGATCTCTGAGTGGT 1695  
Qy 1079 GGGTATCCAGTTTGTCAAACTGATCGCTGCAACCTTGGCTGGAAGGGAATAG 1138  
Db 1696 GGGTCTCTCAGTTTGTCAAACTGATCGCTGCAACCTTGGCTGGAAGGGAATAG 1755  
Qy 1139 AAGAACCACCAAGAGCTTGGCTTCAAACTGATCGCTGCAACCTTGGCTGGAAGGGAATAG 1198  
Db 1756 AAGAACCACCAAGAGCTTGGCTTCAAACTGATCGCTGCAACCTTGGCTGGAAGGGAATAG 1815  
Qy 1199 CTGCAAAAGTGGGAACAGAGCAGCTTCCATCCATTCAGGAATPACATGTCACAGTTG 1258  
Db 1816 CAGCAAAAGTGGGAACAGAGCAGCTTCCATCCATTCAGGAATPACATGTCACAGTTG 1875  
Qy 1259 AAGAATTTTTCAGCTTCTGCAACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1318  
Db 1876 AAGAATTTTTCAGCTTCTGCAACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1935  
Qy 1319 CCACTGATGACCTTCTTTGTTTAAAGGAGGCAAGACAAAGTACTTCCAAATATGAATTTA 1378  
Db 1936 CCAAGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACTTCCAAATATGAATTTA 1995  
Qy 1379 TTAGTGATAACTCTATTTCTTGGTCACTGAGTCACTACAAACGATPACACAGAAATTCAC 1438  
Db 1996 TTAGTGATAACTCTATTTCTTGGTCACTGAGTCACTACAAACGATPACACAGAAATTCAC 2055  
Qy 1439 TTCGGGGCGTATCTCGGATPACACTTTCTCTCCAGGCTGACTTCTTGTGTGTACTT 1498  
Db 2056 TTCGTGGAGTGATCTCGGATPACACTTTCTCTCCAGGAGCTTCTTGTGTGTACTT 2115  
Qy 1499 TTTTCATCCAGCTCTGAGGTTGCTTATGAATCATGCAACATGCTGATCTGATGCT 1558  
Db 2116 TTTTCATCCAGCTCTGAGGTTGCTTATGAATCATGCAACATGCTGATCTGATGCT 2175  
Qy 1559 CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACACAC 1618  
Db 2176 CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACACATC 2235  
Qy 1619 AGATTGCAATTTATCTCACCACCTCGAACTGAAGAGGAAATCCCATGGAACCTGGAG 1678  
Db 2236 AAATTTGCCATTTATGCTCACCACCCGAACTGCAATGAAATTTCCCATGGAACCTGGAG 2295  
Qy 1679 ATATCATTTGGTGTGCTGGAACCACTTGAATGTTTACTCTAAAGGCTGCAACAGAAAC 1738  
Db 2296 ATATCATTTGGTGTGCTGGAATCAATGGAATGCTTCTTAAAGGCTGCAACAGAAAT 2355  
Qy 1739 TAGGAAACAGGCTGTACCTTCTTCAAGTCCGAGAGAGATGAAGAACAGTCAAT 1798  
Db 2356 TGGGAAGACAGGCTGTATATCTCTTCAAGTCCGAGAGAGATGAAGAACAGTCAAT 2415  
Qy 1799 ACCCTACATATCTGAAAGCTGAAATATAGAGTGAAGTGAAGATTA----- 1848

Db 2416 ACCCCACATATCTCTGAGGCTGAGAAATAAAGCTCAGATGGAAGATATAAACGACCAACT 2475  
Qy 1849 -----CAACAGAAATTTAGTTTCAGACCATCTCAGCCAGCAGAGACACCCAG----- 1893  
Db 2476 CAGTTTCACCAAACTCAGTTTCAAACTTTGAGCCAACTGTAGATGAAGAGGCTCTGA 2535  
Qy 1894 -----ACTAACATATGTTTTCATGACACATGCTCCCAACCAAGAGCAAGTGGGAACCT 1949  
Db 2536 TCTAACAAAATAAGTTTATATAGTAGATATCTCTCAGCACCACCAAGAGCAGCTGGGAACCTGA 2595  
Qy 1950 CAGATGCTGC-ACTGGTGGAGCGCTCTTTGTGAGGCTGCTGTGCGCTCAAGCC 2004  
Db 2596 CATAGGCTTCAATTTGGTGGATTCCTCTTTAAACAGGCGCTGCAATGCTCATACCC 2651

RESULT 10  
HSY17979  
LOCUS HSY17979 3280 bp mRNA linear PRI 06-NOV-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase  
transcript A2.  
ACCESSION Y17979  
VERSION Y17979.1 GI:3451268  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Cailleau, A., Balanzino, L., Candelier, J.J., Oriol, R. and  
Mollicone, R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3280)  
AUTHORS Cailleau, A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V.  
Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
Location/Qualifiers  
1..3280  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/clone="FUT8 transcript A2"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..3280  
/gene="FUT8"  
1..389  
/gene="FUT8"  
/note="A"  
390..488  
/gene="FUT8"  
/note="C"  
489..>2444  
/gene="FUT8"  
/note="D"  
717..2444  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76988.1"  
/db\_xref="GI:3451269"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPWTGSWRWIMLILFAWGTLLFYIGHVLVRDNDHPDHSRELS  
KILAKLERLQRQNEELRRMAESLRIPEGIDQGPAGIRVRVLEQLVKAKEQIENYKK  
QTRNGLKDEILRRRIENGAKELPFIQSELKKLKNLEGNLORHAFELLDGHHH  
RSIMDLXYLSOTDGDWEKEAKDITELVORRI TYLQNPKDCSKAKLVNCNKGC  
GYGQLHHVYCFMIAIYGTORTLILESQNWRYATGWETVFRPVSCTCTDRSGISTGH

WSGVBKDNQVVELPIVDLSLHPPPYLPLAVPDLADRLVRVHGDPVWVWVQFVKY  
LIRPOMLEKIEBATKLGPKHPVIGVHVRTDKVCTEAFHP IEBYVHVBEHFOI  
LREPVNDKXVLYATDDPSLLKEAKTKYVNYEFISONSISWSAGLHNRVTENSLRGV  
ILDHFLSQADFLVCTESSQVCRVAYEIMQTLHPDASANFHSLLDDIYFGQVQHNOI  
AIYAHQRTADEIPMEPDIIIGVAGNHWGDSKGVNRKLGRTGLYPSIKVREKIEYVK  
YPTYPEAEK

## variation

2506

/gene="FUT8"  
/note="I1"

/frequency="5"

/replace="c"

3046 .3051

/gene="FUT8"

/note="putative"

3058 .3063

/gene="FUT8"

/note="putative"

3157 .3162

/gene="FUT8"

/note="putative"

BASE COUNT 934 a 718 c 743 g 880 t 5 others

## ORIGIN

Query Match 78.4%; Score 1575.2; DB 9; Length 3280;

Best Local Similarity 87.5%; Pred. No. 0;

Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;

Qy	1	AACGAACCTTATTTCTGTGTGGCTAACTAGAACCGAGTACAAATGTTTCCAAATCTT	60
Db	616	AACGAAGCTCTTACCTGTGCACTAACTAGAACCGAGTACAAATGTTTCCAAATCTT	675
Qy	61	TGAGCTCCGGAAGACA--GAAGGAGTGTGAACCTCTGAAATCGCGCATGGACTGGTT	118
Db	676	TGAGCTCAGAGCTCCAGGGAAGTGAGTTGAANAATCTGAAATCGCGCCATGGACTGGTT	735
Qy	119	CCTGGCTGTGAATATGCTCAATCTTTTGTGCTGGGGACCTTATGTGTTTATATAGGTG	178
Db	736	CCTGGCTGTGAATATGCTCAATCTTTTGTGCTGGGGACCTTGTGTTTATATAGGTG	795
Qy	179	GTCATTTGGTTCGAGATTAATGACCCCTGACCACTTAGCAGAGAACTCCAGATTC	238
Db	796	GTCATTTGGTTCGAGATTAATGACCACTCTGATCACTTAGCCGAGAACTGCCAAGATTC	855
Qy	239	TGCAAGCTCGAGCGCTTAAACACAAATGAAGCTTCAGGAGAGTGCCTGAGTCTC	298
Db	856	TGGCAAGCTTGAACGCTTAAACACAGAGATGAAGCTTGAGCGGAATGGCCGAATCTC	915
Qy	299	TCCGAATACCAAGGCGCTTATGATCAGGGGACAGCTACAGGAAGTCCGCTGTTTAG	358
Db	916	TCCGATACCAAGGCGCTTATGATCAGGGGCGAGCTATAGGAAGTACGCGTTTAG	975
Qy	359	AAGAACAGCTGTTAAGGCCAAAGAACAGATTGAAATTAACAGAAACAGCTAGGAATG	418
Db	976	AAGAGCAGCTGTTAAGGCCAAAGAACAGATTGAAATTAACAGAAACAGACCAAGAAATG	1035
Qy	419	ATCTGGGAAGGATCATGAATCTTAAAGGAGGAGATTGAAATGAGAGTAAAGAGCTCT	478
Db	1036	GTCCTGGGAAGGATCATGAATCTTAAAGGAGGAGATTGAAATGAGAGTAAAGAGCTCT	1095
Qy	479	GTTTCTTCAAAAGTGAATTAAGAAATTAAGAAATTAAGAAAGGAAACGAACTCCAAA	538
Db	1096	GTTTCTTCTACAGATGAATTAAGAAATTAAGAAATTAAGAAAGGAAATGAATCCAAA	1155
Qy	539	GACATGACAGTGAATTTCTTTGGATTTAGGACATCATGAAGGTCTATCATGACAGATC	598
Db	1156	GACATGACAGTGAATTTCTTTGGATTTAGGACATCATGAAGGTCTATATATGCGGATC	1215
Qy	599	TATACTACTCAGTCAACAGATGAGCAGCTGAGTGGCGGGAAGAAAGAACCCAAAGATC	658
Db	1216	TATACTACTCAGTCAACAGATGAGCAGCTGAGTGGCGGGAAGAAAGAACCCAAAGATC	1275
Qy	659	TGACAGAGCTGCTCCAGCGGAGATAACATATCTGCAGAAATCCCAAGGACTGACAGCAAG	718

Db	1276	TGACAGAACTGGTTCAGCGGAGAAATAACATATCTTCAGAAATCCCAAGGACTGCAGCAAAAG	1335
Qy	719	CCAGAAAGCTGGTATGTAATATCAAAAGAGCTGTGGCTATGATGTCAACTCCCATCATG	778
Db	1336	CCAAAGAGCTGGTGTGTAATATCAAAAGAGCTGTGGCTATGCTGTGAGCTTCATCATG	1395
Qy	779	TGGTTTACTGCTTTCATGATTTGCTTATGGCAACCCAGCAACACTCATCTTGGAAATCTCAGA	838
Db	1396	TGGTCTACTCTTCATGATTTGCATATGGCACCCAGCGAAACACTCATCTTGGAAATCTCAGA	1455
Qy	839	ATTGGCGCTATGCTCTGGAGGATGGAGAGCTGTTTATAGACCTGTGATGAGCAGATGCA	898
Db	1456	ATTGGCGCTATGCTCTGGTGGATGGAGAGCTGTTTATAGGCGCTGTAAAGTGAGACATGCA	1515
Qy	899	CAGACAGCTCTGGCTCTCCACTGGACACTGGTCAGGTGAAGTGAAGCAAAAATGTTTC	958
Db	1516	CAGACAGATCTGGCTCTCCACTGGACACTGGTCAGGTGAAGTGAAGCAAAAATGTTTC	1575
Qy	959	AAGTGTGCGAGCTCCCAATTTGATAGACAGCTCCATCTCGTCTCTTACTTACCTTGG	1018
Db	1576	AAGTGTGCGAGCTCCCAATTTGATAGACAGCTCTTCATCCCGTCTCTCCATATTTACCTTGG	1635
Qy	1019	CTGTACAGAAAGACCTTGCAGATCGACTCTCTGAGAGTCCATGTGATCCTGCACTGGGT	1078
Db	1636	CTGTACAGAAAGACCTTGCAGATCGACTCTGTACAGAGTGCATGTGCTGCACTGGGT	1695
Qy	1079	GGGTATCCCAAGTTGTCAAAATACTTGTATCCGTCACAACTTGGCTCGAAAGGAAATAG	1138
Db	1696	GGGTATCCCAAGTTGTCAAAATACTTGTATCCGTCACAACTTGGCTCGAAAGGAAATAG	1755
Qy	1139	AAGAAACACCAAGAAAGCTTGGCTTCAAAATCAAGTATTTGGAGTCCATGTCCAGCGCA	1198
Db	1756	AAGAAACACCAAGAAAGCTTGGCTTCAAAATCAAGTATTTGGAGTCCATGTCCAGCGCA	1815
Qy	1199	CTGACAAAGTGGGAAAGAGAGAGCTTCCATCCATTTGAGGAATATATGATCACAGTTG	1258
Db	1816	CAGACAAAGTGGGAAAGAGAGCTTCCATCCATTTGAGGAATATATGATCACAGTTG	1875
Qy	1259	AAGAAACATTTTCAGCTTCTCGAACGCAAGATGAAGTGGATAAAAAGAGTGTATCTGG	1318
Db	1876	AAGAAACATTTTCAGCTTCTCGAACGCAAGATGAAGTGGATAAAAAGAGTGTATCTGG	1935
Qy	1319	CCATGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA	1378
Db	1936	CCATGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA	1995
Qy	1379	TTAGTGATAACTCTATTTCTTGGTTCAGTGCATACACACCGATACACAGAAATTCAC	1438
Db	1996	TTAGTGATAACTCTATTTCTTGGTTCAGTGCATACACACCGATACACAGAAATTCAC	2055
Qy	1439	TTGGGGGCTGATCTCTGATATACACTTCTCTCCAGGCTGACTTCTCTGTGTGACTT	1498
Db	2056	TTCTGTGAGTATCTCTGATATACATTTCTCTCTCAGGCGAGCTTCTCTGTGTGACTT	2115
Qy	1499	TTTCATCTCCAGGCTGTGAGGCTTGTATGAAATCATGCAAACTCATCTCTGATGCT	1558
Db	2116	TTTCATCTCCAGGCTGTGAGGCTTGTATGAAATCATGCAAACTCATCTCTGATGCT	2175
Qy	1559	CTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCACAAAC	1618
Db	2176	CTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCACAAAC	2235
Qy	1619	AGATTGAGTATCTCTCCCACTCCCACTGAAGAGGAAATCCCATGCAACTGAG	1678
Db	2236	AAATTGCAATTTATGCTTACCACTCCCACTGAGATGAAATTTCCATGCAACTGAG	2295
Qy	1679	ATATCATTTGGTGTGGCTTGAACCACTTGAATGCTTACTCTTAAAGGTGTCAACAGAAAC	1738
Db	2296	ATATCATTTGGTGTGGCTTGAACCACTTGAATGCTTACTCTTAAAGGTGTCAACAGAAAC	2355
Qy	1739	TAGGAAACACAGGCTGTGACCTTCTTCAAAAGTCCGAGAGAGATGAACATGCTCAAT	1798
Db	2356	TGGGAGGACGGGCTTATATCCCTCTTCAAAAGTTCGAGAGAGATGAACATGCTCAAGT	2415



QY 1799 ACCCTACATATCTCTGAAGCTGAAATAAGAGATGAGTGTAGAGATTAA----- 1848  
Db 2416 ACCCCACATATCTCTGAGGCTGAGAAATAAGCTCAGATGGAAGAGATAAACGACCAAACT 2475  
QY 1849 -----CAACAGAAATTAGTTCAGACCATCTTCAGCAAGCAGAGACCCAG----- 1893  
Db 2476 CAGTTCCGACCAAACTCAGTTCAAACTTGGAGCCAACTGTAGATGAAGAGGGCTCTGA 2535  
QY 1894 -----ACTACATATGTTTCAATTCAGACAGATCTCCGACCAAGAGCAAGTGGGAACCT 1949  
Db 2536 TCTAATAAAATAGGTTATAGTAGATATCTCAGCACCAGAGCAGCTGGAACTGA 2595  
QY 1950 CAGATGCTGC-ACTGTGGAAGCGCTCTTTGTGAAGGGCTGTGTCCTCAAGCC 2004  
Db 2596 CATAGGCTTCAATGGTGGAAATCTCTTTTAAACAGGGCTGCAATGCCCTCATACCC 2651

## RESULT 11

HSA536055

LOCUS

DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice variant B5.

ACCESSION

AJ536055

VERSION

AJ536055.1

KEYWORDS

alpha6-fucosyltransferase; FUT8 gene.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Mollicone, R., Michalewski, J.C., Bauvy, C., Cailletau-Thomas, A.,

Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and

Oriol, R.

Splice variants of alpha6-fucosyltransferase are expressed early in

human embryogenesis

Unpublished

REFERENCE

2 (bases 1 to 3568)

Oriol, R.

Direct Submission

Submitted (06-JAN-2003) Oriol R., U504, INSERM, 16 Avenue Paul

Vaillant-Couturier, Villejuif, 94807, FRANCE

related splice variants AJ536053.1, AJ536054.1 and AJ536056.1.

Location/Qualifiers

1..3568

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="14"

/map="14q23.2"

/clone="B2D-L1"

/tissue\_type="whole embryo"

/dev\_stage="40-70 days"

/country="France:Paris"

1..3568

/gene="FUT8"

1098..2825

/gene="FUT8"

/EC\_number="2.4.1.68"

/function="adds fuc in alpha1,6 to the internal GlcNAc of

the core chitobiose"

/codon\_start=1

/evidence=experimental

/product="alpha6-fucosyltransferase"

/protein\_id="CAD59926.1"

/db\_xref="GI:27552400"

/translation="MRPVTGSRWIMLILFAMGTLILFYIGHLVRDNDHPDHSRELS

KILAKLERKQNEDLRMAESLRIPEDIPQGPAGIRVRVLEQLVKAEQIENYKK

QTRNLGKDHDLRRRIENGAKELWFLOSELKKLKNLELQRLHAFELDLGHHE

RSIMTDLYLSQTAGDWREKADLTTELVRRTYVLQNPQDCSKAKLVCNINKC

GYGCOLHVRVCFMAYGTQRTILLESQNRWYATGWTVPFRPSETCDRSGLSTGH

WSGEVKDKVQVDFLPIVDSLHPRPPPLPLAVPDLADLRVVRHGDPAVMWSQFVKY

LIRPQPLWLEKSIIEATKKLGFKHPVIGVHVHVRTDKVGEAAFPPIEYVMVHEHFOL

LARMQVDRKRVYLATDDPSLLKEAKTKYPNYEFISDSISWSAGLHNRHTENSIRGV  
ILDIHFLSQADFLVCTFSQVRYVEIMQTLHPDASANSFHSLLDDIYFFGQNAHQI  
AIYAHQPRTRADEIPMEPGDIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVRKREIETVK  
VPTVPEAEK"  
3538..3543  
/gene="FUT8"  
3553  
/gene="FUT8"  
BASE COUNT 1054 a 751 c 788 g 975 t  
ORIGIN

Query Match 78.4%; Score 1575.2; DB 9; Length 3568;

Best Local Similarity 87.5%; Pred. No. 0;

Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;

QY 1 AACAGAAACTTATTTTCTGTGTGGCTAACTAGAACAGAGTACAAATGTTTCCAAATCTT 60  
Db 997 AACAGAACTTATTTCACTGTGCACTAACTAGAAACAGAGTTACAAATGTTTCAATCTT 1056  
QY 61 TGAGCTCCGAGAGACACA--GAAGGGAGTTGAAAACCTCTGAAAATCGGGCATGGACTGGTT 118  
Db 1057 TGAGCTCCAGGACTCCAGGGAAGTGAATGAAAATCTGAAAATCGGGCATGGACTGGTT 1116  
QY 119 CTTGGCGTTGATATGCTCACTTTTTCCTGGGGAGCTTATTTGTTTATATAGTG 178  
Db 1117 CTTGGCGTTGATATGCTCACTTTTTCCTGGGGAGCTTATTTGTTTATATAGTG 1176  
QY 179 GTCAATTTGTTTCGAGATAATGACCACCTGACCACTTCTAGCAGAGAACTCTCCAAAGATTC 238  
Db 1177 GTCAATTTGTTTCGAGATAATGACCACCTGACCACTTCTAGCAGAGAACTCTCCAAAGATTC 1236  
QY 239 TTGCAAAAGCTGGAGCGCTTAAACCAACAAATTAAGACTTTGAGGAGAAATGGCTGAGTCTC 298  
Db 1237 TGCAAAAGCTTGAACGCTTAAACCAAGAGTGAAGACTTTGAGGCGAAATGGCCGAAATCTC 1296  
QY 299 TCCGAATACCAAGAGGCGCTTATGATCAGGGGAGCTCAGGAGAGAGTCCGCTGTTTAG 358  
Db 1297 TCCGATACCAAGAGGCGCTTATGATCAGGGGAGCTTATGAGGAGAGTATGCGCTTTTAG 1356  
QY 359 AAGAACAGCTTGTAAAGCCAAAGAACAGATTTGAAAATTTACAAGAAACAAGCTAGGAATG 418  
Db 1357 AAGAGAGCTTGTAAAGCCAAAGAACAGATTTGAAAATTTACAAGAAACAAGCTAGGAATG 1416  
QY 419 ATCTGGAAAGAGTATCATGAAATCTTAAAGGAGGAGTTGAAAATTTGAGAGCTCTCT 478  
Db 1417 GTCTGGGGAAGGATCATGAAATCTCTGAGGAGGAGTTGAAAATTTGAGAGCTCTCT 1476  
QY 479 GGTGTTTTCTACAAAGTGAATTTGAAGAAATTTAGAGAAATAGAGGAAACGAACTCCAA 538  
Db 1477 GGTGTTTTCTACAGAGTGAATTTGAAGAAATTTAAAGAACTTTAGAGGAAATGAACTCCAA 1536  
QY 539 GACATGAGATGAAATTTCTTTTGGATTTAGACATCATGAAAGGTCTATCATGACAGATC 598  
Db 1537 GACATGAGATGAAATTTCTTTTGGATTTAGACATCATGAAAGGTCTATATATGAGGATC 1596  
QY 599 TATACTACCTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGGAAGAAAGAGCCAAAGATC 658  
Db 1597 TATACTACCTCAGTCAAGAGATGGAGCAGGTGATTTGCGGGGAAAGAGGCGCAAGATC 1656  
QY 659 TGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCGAGAAATCCCAAGAGCTGCGAGCAAG 718  
Db 1657 TGACAGAACTGGTTTCAGCGGAGAAATACATATCTTCAGAAATCCCAAGAGCTGCGAGCAAG 1716  
QY 719 CCAGAAAGCTGGTGTATTAATCAACAAAGGCTGTGGCTATGAGTGTCAACTCCATCATG 778  
Db 1717 CCAGAAAGCTGGTGTATTAATCAACAAAGGCTGTGGCTATGAGTGTCAACTCCATCATG 1776  
QY 779 TGGTTACTGCTTCATGATTTGCTTATGGCACCAGCGAAACACTCATCTTGAATCTCAGA 838  
Db 1777 TGGTCTACTGCTTCATGATTTGCATATGGCACCAGCGAAACACTCATCTTGAATCTCAGA 1836  
QY 839 ATTGGCGCTATGCTACTTGGAGGATGGAGAGCTGTTTATAGACCTGTAAAGTGAAGCATGCA 898





Best Local Similarity 87.5%; Pred. No. 0; Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;			
QY	1	AACAGAACTTATTTTCCTGCTGCTGCTAATAGAACACAGAGTACAAATGTTTCCAAATCTT	60
Db	1095	AACAGAACTTATTTTCCTGCTGCTGCTAATAGAACACAGAGTACAAATGTTTCCAAATCTT	1154
QY	61	TGAGCTCCGAGAGACA--GAAGGAGTGTGAACACTCTGAAATGCGGGCAGTGGCT	118
Db	1155	TGAGCTCCAGGACTCCAGGGAAGTGAGTTGAAATCTGAAATGCGGCATGAGCTGCT	1214
QY	119	CTGGGCTTGGATATATGCTCATCTTTTCCCTGGGGACCTTATTTGTTATATAGTG	178
Db	1215	CCTGGCTTGGATATATGCTCATCTTTTCCCTGGGGACCTTGTGTTTATATAGTG	1274
QY	179	GTCAATTTGGTTCGAGATATGACCACTGACCATTTCTAGCAGAGAACTCTCCAGATTC	238
Db	1275	GTCACTTGTAGAGATATGACCACTGACCATTTCTAGCAGAGAACTCTCCAGATTC	1334
QY	239	TTGCAAGCTGGAGCGCTTTAAACCAACAAATGAAGACTTGAGGAGAAATGGCTGAGTCT	298
Db	1335	TGGCAAGCTTGAACGCTTAAACCAACAGCAATGAAGACTTGAGGAGAAATGGCTGAGTCT	1394
QY	299	TCCGAATACCAAGGCCCTATTGATCAGGGGACAGCTACAGGAGAGAGTCCGTTTGTAG	358
Db	1395	TCCGGATACCAAGGCCCTATTGATCAGGGGACAGCTATAGGAGAGAGTACGCGTGTAG	1454
QY	359	AGACAGCTGTTTAAAGCCAAAGAACAGATTGAAATTTACAAGAAACAAGCTAGGAATG	418
Db	1455	AAGACAGCTGTTTAAAGCCAAAGAACAGATTGAAATTTACAAGAAACAAGCTAGGAATG	1514
QY	419	ATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTTGAAATGGAGCTTAAAGAGCTCT	478
Db	1515	GTCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTTGAAATGGAGCTTAAAGAGCTCT	1574
QY	479	GGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAGAAATTTAAGAGAAACGAATCTCAAA	538
Db	1575	GGTTTTTTCTACAGAGTGAATTTGAAGAAATTTAAGAAATTTAAGAGAAATGAACTCCAAA	1634
QY	539	GACATGCAGATGAATCTTTTGGATTTAGGACATCATGAAAGGCTCATCATGACAGATC	598
Db	1635	GACATGCAGATGAATCTTTTGGATTTAGGACATCATGAAAGGCTCATCATGACGATC	1694
QY	599	TATACTACCTCAGTCAAAACAGATGGAGCAGGTGAGTGGCGGGAAGAAAGCCAAAGATC	658
Db	1695	TATACTACCTCAGTCAAGACAGATGGAGCAGGTGATTTGGCGGGAAGAAAGCCAAAGATC	1754
QY	659	TGACAGAGCTGGTCCAGCGGAGAAATAATATCTGCAGAAATCCCAAGGACTGCGAGCAAG	718
Db	1755	TGACAGAACTGGTTCAGCGGAGAAATAATATCTTCAGAAATCCCAAGGACTGCGAGCAAG	1814
QY	719	CCAGAAAGCTGTATGTAATATCAACAAGGCTGTGGCTATGGATGTCAACTCCATCATG	778
Db	1815	CCAAAAGCTGTGTATATATCAACAAGGCTGTGGCTATGGCTGTGAGCTCCATCATG	1874
QY	779	TGGTTTACTGCTTCAATGATTTGCTTATGGCACCCAGCGAACACTCATCTTGGAAATCTCAGA	838
Db	1875	TGGTCTACTGCTTCAATGATTTGATATGCAATGCAACCCAGCGAACACTCATCTTGGAAATCTCAGA	1934
QY	839	ATTGGCGCTATGCTACTGAGAGATGGGAGACTGTGTTTAGACCTGTAAAGTGAAGATGCA	898
Db	1935	ATTGGCGCTATGCTACTGCTGAGATGGGAGACTGTATTTAGGCGCTGTAAAGTGAAGATGCA	1994
QY	899	CAGACAGCTTGGCGCTCTCCACTGACACTGTGTAGAGTGAAGTGAAGGACAAATGTTTC	958
Db	1995	CAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAATGTTTC	2054
QY	959	AAGTGGTCCAGCTCCCACTTTAGACAGCTCCATCTCGTCCCTCTTACTTACCTTTGG	1018
Db	2055	AAGTGGTCCAGCTCCCACTTTAGACAGCTCTTCATCCCGCTCCCTCATTTTACCTTTGG	2114
QY	1019	CTGTACCAAGAGACCTTCAGATCGACTCTTCGAGAGTCCATGGTGTGATCCTCGAGTGTGGT	1078

Db	2115	CTGTACCAAGAGACCTCGCAGATCGACTTGTACAGTGCATGGTGACCCCTCGAGTGTGGT	2174
QY	1079	GGGTATCCCAAGTTTGTCAAATACCTTGTATCCGTCCCAACCTTGGCTGGAAAGGAAATAG	1138
Db	2175	GGGTGTCTCAGTTTGTCAAATACCTTGTATCCGCCCCACAGCCTTGGCTAGAAAAGAAATAG	2234
QY	1139	AAGAAAACCAAGAGCTTGGCTTCAAACATCCAGTTATTTGGAGTGCATGTACAGCGCA	1198
Db	2235	AAGAAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTATTTGGAGTGCATGTACAGCGCA	2294
QY	1199	CTGACAAAGTGGGAACAGAGCAGCTTCCATCCATTTGAGGAATACATGTACACGTTG	1258
Db	2295	CAGACAAAGTGGGAACAGAGCAGCTTCCATCCATTTGAGGAATACATGTATGATGTTG	2354
QY	1259	AAGAACAATTTTCAAGCTTCTCGAACGCGAAGTGAAGTGAATGGAATGATGTATCTGG	1318
Db	2355	AAGAACAATTTTCAAGCTTCTCGAACGCGAAGTGAAGTGAATGGAATGATGTATGTTG	2414
QY	1319	CCACTGATGACCTTCTTTGTTAAAGGAGGCAAGACAAAGTACTCCAAATATGAATTTA	1378
Db	2415	CCACAGATGACCTTCTTTTAAAGGAGGCAAGACAAAGTACTCCCAATATGAATTTA	2474
QY	1379	TTAGTGTAACTCTATTTCTTGTTCAGTGCATACACACCGATACACAGAAATTCAC	1438
Db	2475	TTAGTGTAACTCTATTTCTTGTTCAGTGCATACACACCGATACACAGAAATTCAC	2534
QY	1439	TTCCGGGCGTGCATCCTCGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTACTT	1498
Db	2535	TTCTGTGAGTGTACTCTGTATATACATTTCTCTCTCAGGAGACTTCTTGTGTGTACTT	2594
QY	1499	TTTCATCCAGGCTGTAGGTTGCTTATGAATCATCTGCAAAACACTGCTGATGCT	1558
Db	2595	TTTCATCCAGGCTGTAGGTTGCTTATGAATCATCTGCAAAACACTGATCTGATGCT	2654
QY	1559	CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACAACT	1618
Db	2655	CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACAACT	2714
QY	1619	AGATTGCAGTTTATCTTCAGCAACCTCGAACTAAGAGGAAATCCCATGGAACCTGGAG	1678
Db	2715	AAATTGCCATTTATGCTCACCACCCGAACTGCAGATGAAATTTCCCATGGAACCTGGAG	2774
QY	1679	ATATCATTTGTTGGCTGGAAACCATTTGGAATGTTACTCTAAAGGTTCTCAACAGAAAC	1738
Db	2775	ATATCATTTGTTGGCTGGAAACCATTTGGAATGTTACTCTAAAGGTTCTCAACAGAAAC	2834
QY	1739	TAGAAAACAGGCTGTACCTTCCCTACAAAGTCCGAGAGAGATAGAAACAGTCAAT	1798
Db	2835	TGGGAAGGACGGGCTTATATCCCTTCTACAAAGTTCGAGAGAGATAGAAACGGTCAAGT	2894
QY	1799	ACCTTACATATCTCTGAAGCTGAAAAATAGAGATGGAGTGAAGAGATTAA-----	1848
Db	2895	ACCCCAATATCTCTGAGGCTGAGAAATTAAGCTCAGATGGAAGAGATAACGACCAACT	2954
QY	1849	-----CAACAGATTTAGTTCAGACCATCTCAGCCCAAGCAGAGAACCCAG-----	1893
Db	2955	CAGTTCGACCAAACTCAGTTTCAAACTTTGAGCAAACTGTAGTAGAAGGGCTCTGA	3014
QY	1894	-----ACTAACATATGTTTCAATGACAGATGCTCCGACCAAGAGCAAGTGGGAACCT	1949
Db	3015	TCTAACAAAATAGGTTTATGATGATGATGATCTCTCAGACCAAGAGAGCTGGGAACTGA	3074
QY	1950	CAGATGCTGC-ACCTGTTGGAAACGCTCTTTTGTGAAGGGCTGCTGTGCCCTCAAGCC	2004
Db	3075	CATAGGCTTCAATTTGGTGAATTCCTCTTTAAAGGGCTGCAATGCTCATACCC	3130

RESULT 13

HSA536054

LOCUS

DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice variant B4.

ACCESSION

AJ536054

3772 bp

mRNA

linear

PRI 08-JAN-2003

VERSION	AJ536054.1	GI:27552397	
KEYWORDS	alpha6-fucosyltransferase; FUT8 gene.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 Mollicone, R., Michalski, J.C., Bauvy, C., Caillet, Thomas, A., Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.		
TITLE	Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3772)		
AUTHORS	Oriol, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-JAN-2003) Oriol R., U504, INSERM, 16 Avenue Paul Vaillant-Couturier, Villejuif, 94807, FRANCE		
COMMENT	related splice variants AJ536053.1, AJ536055.1 and AJ536056.1.		
FEATURES	Location/Qualifiers		
source	1..3772		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="14"		
	/map="14q23.2"		
	/clone="B2CD-L2"		
	/tissue_type="whole embryo"		
	/dev_stages="40-70 days"		
	/country="France:Paris"		
gene	1..3772		
	/gene="FUT8"		
CDS	1196..2923		
	/gene="FUT8"		
	/EC_number="2.4.1.68"		
	/function="adds fuc in alpha1,6 to the internal GlcNAc of the core chitobiose"		
	/codon_start=1		
	/evidence=experimental		
	/product="alpha6-fucosyltransferase"		
	/protein_id="CAD59925.1"		
	/db_xref="GI:27552398"		
	/translation="MRPWTGSRWIMLILFAWGLTLFVIGHLVRDNDPHDSRELS KILAKRLKQNDLRMAESLRIPEGIDQGAIGRVRLBOLVKAKEQIENYKK QTRNGLKQKHILRRLENGAKELWFLQSELKLNLELORHAEFLDLGHE RSMIDLYLSOTDAGDWEKEADLHLELQRIYLYQNPCKSKAKLVNINKGC GYGCLHHVYVCFMIAYGQRTILILESONNRYATGWTVPVRSCTDRSGISTGH WSGVKQKNVQVLELPVDSLHPHPYPLAVPEDLADRLVRHGDPAVMVQSFVK LIRPQPLEKEIEBATKLGPKHPVIGVHVVRTDKVTEAFHPHIEYMHVIEHFOL LARMOVDKRVLYATDPSLLKEAKTYPNVEFISNLSWSAGLHNRVTENSLRGV ILDIHELQADFLVCTPSSQVCRVAYEIMQTLHPDASANPHSLDDIYVFGONAHNOI AIYAHQPTADEIEMPEPDIILGVAGNHWGYSKGNRLKRGTLGILYPSYKVKREKLETVK YPTYPEAEK"		
polyA_signal	3723..3728		
	/gene="FUT8"		
polyA_site	3757		
	/gene="FUT8"		
BASE COUNT	1110 a 777 c 830 g 1055 t		
ORIGIN			
Query Match	78.4%;	Score 1575.2;	DB 9; Length 3772;
Best Local Similarity	87.5%;	Pred. No. 0;	
Matches 1781;	Conservative 0;	Mismatches 223;	Indels 32; Gaps 4;
Qy	1 AACAGAACTTATTTCTGTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 60		
Db	1095 AACAGAACTTATTTCTGTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 1154		
Qy	61 TGAGCTCCGAGAGACA--GAAGGAGTTGAAACTCTGAAAATCGGGCATGGAGTGGTT 118		
Db	1155 TGAGCTCCGAGAGTCTCAGGAGGAGTGGTTGAAAATCTGAAAATCGGGCATGGAGTGGTT 1214		
Qy	119 CCTGGCGTTGATTATGCTATCTTTTGGCTGGGGACCTTATTTGTTTATATAGTGT 178		
1215	CTTGGCGTTGGAATATGCTCATTTCTTTTGGCTGGGGACCTTGGCTCTTTTATATAGTGT 1274	Db	
179	GTCAATTTGGTTCCAGATAATGACCCACCTGACCATTTCTAGCAGAGAACTCTCCAGATTC 238	Qy	
1275	GTCACTTTGGTACAGATAATGACCATCTCTGATCACTCTAGCCGAGAACTGTCCAGATTC 1334	Db	
239	TTGCAAAAGCTGGAGCGCTTAAAAACAACAAATGAAGACTTTGAGGAGAAATGGCTGAGTCT 298	Qy	
1335	TGGCAAAAGCTTGAACGCTTAAACACAGAGATGAAGACTTTGAGGCGAATGGCCGAATCTC 1394	Db	
299	TCCGAATACAGAGGCGCTTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTGTTTATAG 358	Qy	
1395	TCCGATACAGAGGCGCTTATTGATCAGGGGCGAGCTATAGGAAGAGTACGCGTTTATAG 1454	Db	
359	AAGAACAGCTTGTAAAGGCCAAAGACAGATTGAATAATTACAAGAAACAGCTAGGAATG 418	Qy	
1455	AAGAGAGCTTGTAAAGGCCAAAGACAGATTGAATAATTACAAGAAACAGACCAAGAAATG 1514	Db	
419	ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGGATTGAAAATGGAGCTTAAAGAGCTCT 478	Qy	
1515	GTCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGGATTGAAAATGGAGCTTAAAGAGCTCT 1574	Db	
479	GGTTTTTCTACAAAGTGAATTTGAAGAAATTAAAGAAATTAGAAAGAAACGAACTCCAAA 538	Qy	
1575	GGTTTTTCTACAGAGTGAATTTGAAGAAATTAAAGAAACTTTAGAAAGAAATGAACTCCAAA 1634	Db	
539	GACATGAGAGTGAATTTCTTTGGATTAGACATCATGAAGGTCTTATCATGACATC 598	Qy	
1635	GACATGAGAGTGAATTTCTTTGGATTAGACATCATGAAGGTCTTATATATGACGATC 1694	Db	
599	TATACTACCTCAGTCAAAACAGATGGAGCAGTGTAGTCGGCGGAAAAAGAACCCAAAAGATC 658	Qy	
1695	TATACTACCTCAGTCAACAGATGGAGCAGTGTATTCGGCGGAAAAAGAACCCAAAAGATC 1754	Db	
659	TGACAGAGCTGGTCCAGCGGAGATAACATATCTGCAGAAATCCCAAGAGCTGCAGCAAAAG 718	Qy	
1755	TGACAGAACTGGTTTTCAGCGGAGATAACATATCTTCAGAAATCCCAAGAGCTGCAGCAAAAG 1814	Db	
719	CCAGAAAGCTGGTATTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG 778	Qy	
1815	CCAAAAAGCTGGTGTGTAATATCAACAAAGGCTGTGGCTATGGCTGTGAGCTCCATCATG 1874	Db	
779	TGGTTTACTGCTTCAVCAATTTGCTTATGGCCACCCAGCAACACTCATCTTGGAAATCTCAGA 838	Qy	
1875	TGGTCTACTGCTTCAVCAATTTGCTTATGGCCACCCAGCAACACTCATCTTGGAAATCTCAGA 1934	Db	
839	ATTGGCGCTATGCTTACTTGGAGGATGGAGACTGTGTTTATAGACCTGTAAAGTGAACATGCA 898	Qy	
1935	ATTGGCGCTATGCTTACTTGGTGGATGGAGACTGTATTTAGGCGCTGTAAAGTGAACATGCA 1994	Db	
899	CAGACAGCTGGGCTTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAATGTTTC 958	Qy	
1995	CAGACAGCTGGGCTTCCACTGGACACTGGTCAGGTGAAGTGAAGGACAAAATGTTTC 2054	Db	
959	AAGTGTGCGAGCTCCCATTTGTAGACAGCTCCATCTCGTCCCTCTTACTTACCCCTGG 1018	Qy	
2055	AAGTGTGCGAGCTTCCCATTTGTAGACAGCTTCTATCCCGCTCTCCATATTTACCCCTGG 2114	Db	
1019	CTGTACCAAGAGACTTTCAGATCGACTCTCTGAGAGTCCATGTGTATCTTGCAGTGGGT 1078	Qy	
2115	CTGTACCAAGAGACTTTCAGATCGACTTGTACAGTGCATGTGTGACCTGCAGTGGGT 2174	Db	
1079	GGGTATCCCGATTTGTCAAAATATTGATCCGTCACAAACCTTGGCTGGAAAGGAAATAG 1138	Qy	
2175	GGGTGTCTCATTTGTCAAAATATTGATCCGCTCCACAGCTTGGCTAGAAAAGAAATAG 2234	Db	
1139	AAGAAACCAACAGAAAGTTGGCTTCAACATCCAGTTATTGGAGTCCATGTCCAGACGCA 1198	Qy	
2235	AAGAAACCAACAGAAAGTTGGCTTCAACATCCAGTTATTGGAGTCCATGTCCAGACGCA 2294	Db	
1199	CTGACAAAGTGGGAAACAGAAAGCAGCTTCCATCCCTATTGAGGAAATACATGTCACCGTTG 1258	Qy	

Db 2295 CAGACAAAGTGGGACAGAAAGCTGCCTTCCATCCCATTTGAAGAGTACATGGTGCATGTTG 2354  
Qy 1259 AAGAAATTTTTCAGCTTCTCGAACGAGCAATGAAAGTGGATAAAAAAGAGAGTGTATCTGG 1318  
Db 2355 AAGAAATTTTTCAGCTTCTCGAACGAGCAATGAAAGTGGATAAAAAAGAGAGTGTATTTGG 2414  
Qy 1319 CACTGATGACCTCTTCTTTGTTAAAGGAGGAAAGACAAAGTACTCCAAATATGAAATTA 1378  
Db 2415 CCACAGATGACCTCTTCTTTATTAAGGAGGAAAGACAAAGTACTCCAAATATGAAATTA 2474  
Qy 1379 TTAGTGATAACTCTATTTCTCGTGCAGCTGGACTACAAACGATACACAGAAATTCAC 1438  
Db 2475 TTAGTGATAACTCTATTTCTCGTGCAGCTGGACTACAAATCGATACACAGAAATTCAC 2534  
Qy 1439 TTGGGGGCGTGATCTCTGGATATACACTTTCTCTCCAGGCTGACATTCCTTGTGTACTT 1498  
Db 2535 TTCTGGGAGTATCTCTGGATATACACTTTCTCTCTCAGGAGACTTCTAGTGTACTT 2594  
Qy 1499 TTTCATCCAGCTCTGTAGGTTGCTTATGAAATCATGCAACACTGATGCTCTGATGCT 1558  
Db 2595 TTTCATCCAGCTCTGTAGGTTGCTTATGAAATCATGCAACACTGATGCTCTGATGCT 2654  
Qy 1559 CTGCAAACTTCCATTTCTAGATGACATCTACTATTTGGAGGCCAAATGCCCAAC 1618  
Db 2655 CTGCAAACTTCCATTTCTAGATGACATCTACTATTTGGAGGCCAGAAATGCCCAATC 2714  
Qy 1619 AGATTGCACTTTATCTCACCACCTCGAACTTAAAGAGGAAATCCCATGGAACCTGGAG 1678  
Db 2715 AAATTGCCATTTATGCTCACCACCTCGAACTTAAAGAGGAAATCCCATGGAACCTGGAG 2774  
Qy 1679 ATATCAATTTGGTGTGGTGAACCAATTTGAATGATGTTACTCTTAAAGTGTCAACAGAAAC 1738  
Db 2775 ATATCAATTTGGTGTGGTGAACCAATTTGAATGATGTTACTCTTAAAGTGTCAACAGAAAT 2834  
Qy 1739 TAGGAAACAGCGCTGTACCTCTCACAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
Db 2835 TGGGAAGGAGCGGCTTATATCTCTCACAAGTTCGAGAGAGATAGAAACAGTCAAT 2894  
Qy 1799 ACCCTACATATCTGAACTGAAATAGAGATGAGTGTAGAGATTAA----- 1848  
Db 2895 ACCCACATATCTGAGCTGAGAAATTAAGCTCAGATGGAAGATAGAAACGACCAACT 2954  
Qy 1849 -----CAACAGATTTAGTTAGACATCTCAGCCAGAGAGACCCAG----- 1893  
Db 2955 CAGTTGCAACCACTCAGTTCAACCAATTTGAGCCAACTGTAGATGAAGGGCTCTGA 3014  
Qy 1894 ----ACTACATATGTTTCATTTGACAGATGCTCCGACCAAGAGCAAGTGGGACCT 1949  
Db 3015 TCTAACAAATTAAGTTATATGATAGATATCTCTCAGCACCAGAGCAAGTGGGAACTGA 3074  
Qy 1950 CAGATGCTGC-ACTGTTGAAGCGCTCTTTTGTGAAGGCTGCTGTGCCCTCAAGCC 2004  
Db 3075 CATAGGCTCAATGTTGGATTCTCTTTTAAAGGGCTGCAATGCTCATACCC 3130

RESULT 14  
HSA536056  
LOCUS HSA536056 4196 bp mRNA linear PRI 08-JAN-2003  
DEFINITION Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice variant B6.  
ACCESSION AJ536056  
VERSION AJ536056.1 GI:27552401  
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Mollicone, R., Michalek, J., C., Bauvy, C., Caillaud-Thomas, A., Candelier, J. J., Martinez-Duncker, I., Breton, C., Codogno, P., and Oriol, R.  
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis

Unpublished  
2 (bases 1 to 4196)  
REFERENCE  
ORiol, R.  
Direct Submission  
Submitted (06-JAN-2003) Oriol, R., U504, INSERM, 16 Avenue Paul  
Vaillant-Couturier, Villejuif, 94807, FRANCE  
related splice variante AJ536053.1, AJ536054.1 and AJ536055.1.  
COMMENT  
FEATURES  
Location/Qualifiers  
1..4196  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23.2"  
/clone="B3CD-L1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
/country="France:Paris"  
1..4196  
/gene="FUT8"  
1726..3453  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/function="adds fuc in alpha1,6 to the internal GlcNAc of the core chitobiose"  
/evidence="experimental"  
/product="alpha6-fucosyltransferase"  
/protein\_id="CAD59927.1"  
/db\_xref="GI:27552402"  
/translation="MRPWTGSWMILIFAWGTLLEYIGGHLVRDNDHPDHSRELS  
KILAKLERQKQEDLRMAESLRIPGPDQPAIGRVRLEEQLVKAKEIENYKK  
QTNRGKDEHILRRRIENGAKELWPFQSEKLKLNLEGNLQHRHDEFLLDLGHE  
RSIMTDLYSQTDGAGDREKAKDLITELVQRRITLYLQNPDKCSAKKLVCNINKG  
GSGQLHHVYCFMIAVGTORTLILSONWRYATGWTVPVSETCTDORSIGSTGH  
WSEVKDKNVQVELPIVDSLHPRPVPLAIPEDLADLRVVRHGDPAVWVSQFVKY  
LIRPQWLEKEIEBATKLGFKHPVIGVHVRDVKVGTAAAPHPHIEYVWVBEHFQL  
LAARMQVKKRVYATDPSLKEAKTKYNYEFISDINSWSAGLHNRVTENLRGV  
ILDIHFLSQADFLVCTFSQCRVAYEIMQTLHPDASANFHSLLDDIYFQGGQNAHQI  
AIYAHOPRTADEIPMEPGDIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVKETVK  
YPTVPEAEK"  
4166..4171  
/gene="FUT8"  
polyA\_site  
4181  
/gene="FUT8"  
BASE COUNT 1245 a 854 c 900 g 1197 t  
ORIGIN  
Query Match 78.4%; Score 1575.2; DB 9; Length 4196;  
Best Local Similarity 87.5%; Pred. No. 0;  
Matches 1781; Conservative 0; Mismatches 223; Indels 32; Gaps 4;  
Qy 1 AACAGAACTTATTTCTCTGTGGCTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 60  
Db 1625 AACAGAACTTATTCACCTGTGCACCTAAGAACAGAGATTACATGTTTTCANATCTT 1684  
Qy 61 TGAGCTCCGAGAGACACA--GAAGGAGGTGAAACTCTGAAAAATCGCGGCACTGGACTGTT 118  
Db 1685 TGAGCTCCAGACTCCAGGGAAGTGAAGTTCGAAATCTGAAAAATCGCGGCACTGGACTGTT 1744  
Qy 119 CTGGCGCTGGATTATGCTCAATCTTTTTCCTGGGGGACCTTATGTTTTATATAGGTG 178  
Db 1745 CTGGCGCTGGATTATGCTCAATCTTTTTCCTGGGGGACCTTCTGCTGTTTTATATAGGTG 1804  
Qy 179 GTCAATTTGTTTCAGATAATGACCCCTGACCATCTACAGAGAACTCTCCAAAGATTC 238  
Db 1805 GTCACTTGGTACAGATAATGACCATCTCTATCTACCTAGCCGAGAACTGTCAGAGATTC 1864  
Qy 239 TTGCAAGCTGGAGCGCTTAAAAACAACAAATGAAGACTTTGAGAGAAATCGGTGAGTCTC 298  
Db 1865 TGGCAAGCTTGAACGCTTAAAAACAGCAGAAATGAAGACTTTGAGCGAAATGCGCGAAATCTC 1924  
Qy 299 TCGAATACAGAAAGGCCCTATTGATCAGGGGACAGCTACAGGAGAGTCCGTGTTTAG 358

1925 TCCGATACAGAGGCCCTATTGATCAGGCGCCAGCTATAGGAAGTAGTACGCGTTTAG 1984  
359 AAGAACAGCTGTTTAAAGCCCAAAGAAAGAGATTGAAATTTACAAAGAACAGCTAGGAATG 418  
1985 AAGAGCAGCTGTTTAAAGCCCAAAGAAAGAGATTGAAATTTACAAAGAACAGACCAAGAAATG 2044  
419 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGAGATTGAAATTTGAGAGCTTAAAGAGCTCT 478  
2045 GTCTGGGAAAGGATCATGAAATCTTAAAGGAGAGATTGAAATTTGAGAGCTTAAAGAGCTCT 2104  
479 GGTATTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTTAGAAAGAAAGAAAGCTCCAAA 538  
2105 GGTATTTTCTACAGATGAATTTGAAGAAATTTAAAGAAATTTAGAAAGAAATTTAGAAAGCTCCAAA 2164  
539 GACATGACAGTGAATTTCTTTTGAATTTAGACATCATGAAAGTCTCATCATGACAGATC 598  
2165 GACATGACAGTGAATTTCTTTTGAATTTAGACATCATGAAAGTCTCATCATGACGATC 2224  
599 TATACCTCAGTCAACAGATGAGCAGCTGAGTGGCGGAAAGAAAGAAAGCTCCAAAGATC 658  
2225 TATACCTCAGTCAACAGATGAGCAGCTGAGTGGCGGAAAGAAAGAAAGCTCCAAAGATC 2284  
659 TGACAGAGCTGGTCCAGCGGAGATTAACATATCTGCAGAAATCCCAAGGACTGCGAGCAAG 718  
2285 TGACAGAGCTGGTCCAGCGGAGATTAACATATCTGCAGAAATCCCAAGGACTGCGAGCAAG 2344  
719 CCAGAAAGCTGGTATTAATTAACAAAGAGCTGTGGCTATGGATGTCACTCATCATG 778  
2345 CCAGAAAGCTGGTATTAATTAACAAAGAGCTGTGGCTATGGATGTCACTCATCATG 2404  
779 TGGTTACTGCTCATGATCTTATGCGCAAGCTGAGTGGCGGAAAGAAAGAAAGCTCCAG 838  
2405 TGGTTACTGCTCATGATCTTATGCGCAAGCTGAGTGGCGGAAAGAAAGAAAGCTCCAG 2464  
839 ATTGGCGTATGCTACTGAGAGTGGGAGACTGTGTTTGAAGCTGTTAGTGAAGTGAATGCA 898  
2465 ATTGGCGTATGCTACTGAGTGGGAGACTGTGTTTGAAGCTGTTAGTGAAGTGAATGCA 2524  
899 CAGACAGTCTGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTTC 958  
2525 CAGACAGTCTGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTTC 2584  
959 AAGTGGTCCAGCTCCCAATGTTAGACAGCTCCATCTCGTCCCTTACTTACCTTGG 1018  
2585 AAGTGGTCCAGCTCCCAATGTTAGACAGCTCCATCTCGTCCCTTACTTACCTTGG 2644  
1019 CTGTACCAAGAGACCTTCAGATCGACTCCTGAGAGTCCATGGTGGTATCCTGCGAGTGGT 1078  
2645 CTGTACCAAGAGACCTTCAGATCGACTCCTGAGAGTCCATGGTGGTATCCTGCGAGTGGT 2704  
1079 GGGTATCCAGTGTTCATAATCTTGAATCCGTCACAACTTGGCTGGAAGGAAATAG 1138  
2705 GGGTGTCTCAGTGTTCATAATCTTGAATCCGTCACAACTTGGCTGGAAGGAAATAG 2764  
1139 AAGAAACCAAGAGAGCTTGGCTTCAACATCGAGTTATTGGAGTCCATCTCAGACGCA 1198  
2765 AAGAAACCAAGAGAGCTTGGCTTCAACATCGAGTTATTGGAGTCCATCTCAGACGCA 2824  
1199 CTGACAAAGTGGGAAAGAGAGCTTCCATCCCATTTGAGGAATACATGGTGTACAGCTTG 1258  
2825 CAGACAAAGTGGGAAAGAGAGCTTCCATCCCATTTGAGGAATACATGGTGTACAGCTTG 2884  
1259 AAGAAATTTTTCAGCTTCTCGAACCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTTCGG 1318  
2885 AAGAAATTTTTCAGCTTCTCGAACCCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTTCGG 2944  
1319 CCACTGATGACCTTCTTTTGAAGGAGGCAAGACAAAGTACTCCAAATTTATGAATTTA 1378  
2945 CCACTGATGACCTTCTTTTGAAGGAGGCAAGACAAAGTACTCCAAATTTATGAATTTA 3004  
1379 TTAGTGATAACTTATTTCTGGTCCAGCTGAGTACACAAAGCTGACAGAAATTTAC 1438

3005 TTAGTGATAACTCTATTCTTCTGGTCCAGTGGAGTGCACAAATCGATACACAGAAATTTAC 3064  
1439 TTGGGGCGGTGATCTGAGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTAATTT 1498  
3065 TTCTGGAGTGTCTGGATATATATTTCTCTCTAGGAGAGTCTTCTAGTGTGTAATTT 3124  
1499 TTTTCATCCCAAGGTCTGAGGTTGCTTATGAAATCATGCAAAACATGCTCATCTGATGCTT 1558  
3125 TTTTCATCCCAAGGTCTGAGGTTGCTTATGAAATCATGCAAAACATGCTCATCTGATGCTT 3184  
1559 CTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCCACAACT 1618  
3185 CTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGGCGCAGAAATGCCCACAACT 3244  
1619 AGATTGAGTCTTATCTCACCACCTCGAACTTAAAGAGGAAATCCCATGGAACCTGGAG 1678  
3245 AATTTGCCATTTATCTCACCACCTCGAACTTAAAGAGTGAATTTCCATGGAACCTGGAG 3304  
1679 ATATCATTTGGTGTCTGGAACCATTTGGAATGTTACTCTTAAAGGTGTCAACAGAAAT 1738  
3305 ATATCATTTGGTGTCTGGAACCATTTGGAATGTTACTCTTAAAGGTGTCAACAGAAAT 3364  
1739 TAGGAAACACAGGCTGTGACCTTCTCAAAAGTCCGAGAGAGATAGAAACATGCTCAAT 1798  
3365 TGGGAGGACGGGCTATATCTCCTCTCAAAAGTTCGAGAGAGATAGAAACGCTCAAGT 3424  
1799 ACCCTACATATCTCGAAGCTGAAATAGAGATGAGTGAAGAGATTAA----- 1848  
3425 ACCCTACATATCTCGAAGCTGAAATAGAGATGAGTGAAGAGATTAAACGACCAAACT 3484  
1849 -----CAACAGAAATTTAGTTTCAAGCACTCTCAGCAAGCAAGCAAGCAAGCAAG 1893  
3485 CAGTTTCAGCAAACTCAGTTTCAAACTTTCAGCAAGCAAGCAAGCAAGCAAGCTCTGA 3544  
1894 -----ACTAACATATGTTTCAATGACAGATGCTCCGCAACCAAGCAAGCAAGCAAGCTT 1949  
3545 TCTAACAAATAAGATATATAGTAGATATCTCTCAGCAAGCAAGCAAGCAAGCAAGCTTGA 3604  
1950 CAGATGCTGC-CTGCTGGAGAGCTCTTGTGAAGGCTGCTGCTGCTTCAAGCC 2004  
3605 CATAGGCTTCAATTTGGTGAATTCCTCTTTTAAAGGCTGCAATGCTTATATCC 3660

## RESULT 15

E15725 LOCUS Human mRNA for alpha-1,6-fucosyltransferase, complete cds.  
DEFINITION Human mRNA for alpha-1,6-fucosyltransferase, complete cds.  
ACCESSION E15725  
VERSION E15725.1 GI:5710498  
KEYWORDS JP 1998084975-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 2100)  
AUTHORS Taniguchi, N., Uozumi, H. and Yanagida, S.  
TITLE ALPHA-1,6-FUCOSYLTRANSFERASE GENE DERIVED FROM HUMAN  
JOURNAL Patent: JP 1998084975-A 1 07-APR-1998;  
TOYOBO CO LTD  
COMMENT OS Homo sapiens (human)  
PN JP 1998084975-A/1  
PD 07-APR-1998  
PF 17-JUN-1997 JP 1997159692  
PR 22-JUL-1996 JP 96P 192260  
PI TANIGUCHI NAOTYKI, UOZUMI HISAFUMI, YANAGIDA SHUSAKU PC  
C12N15/09, C07H21/04, C12N1/21, C12N9/10, (C12N15/09, C12R1:91), PC  
(C12N1/21).  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FH source 1. .2100  
FT

```
FT      /organism='Homo sapiens'
FT      /cell_type='stomach cancer cell' FT
FT      /cell_lines='MKN45'
FT      198..1925
FT      /product='alpha-1,6-fucosyltransferase'.
FEATURES
  source      Location/Qualifiers
              1..2100
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
BASE COUNT  651 a  436 c  489 g  524 t
ORIGIN
Query Match      78.3%; Score 1572.8; DB 6; Length 2100;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1694; Conservative 0; Mismatches 182; Indels 2; Gaps 1;
QY 1 AACAGAACTTATTTCTGTGTGGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATCTTT 60
Db 1 AACAGAACTTATTTCTGTGTGGCTAACTAGAACCCAGAGTACAAATGTTTCCAAATCTTT 156
QY 61 TGAGCTCCGAGAGACA--GAAGGAGTTGAACTCTGAAATCGGGCATGGAGCTGTT 118
Db 157 TGAGCTCCAGGACTCCAGGGAAGTGAAGTTGAAATCTGAAATCGGGCATGGAGCTGTT 216
QY 119 CTTGGCGTTGGATTATGCTCATCTTTTGGCTGGGGACCTTATGTTTATATAGGTG 178
Db 217 CTTGGCGTTGGATTATGCTCATCTTTTGGCTGGGGACCTTATGTTTATATAGGTG 276
QY 179 GTCAATTTGGTTGAGATAATGACACCTGACCATTTAGCAGAGAACTCTCCAGATTC 238
Db 277 GTCACTTTGGTACGATAATGACCATCTCTGATCACTTAGCCGAGAACTCTCCAGATTC 336
QY 239 TTGCAAACTGAGCGCTTAAACACAAATGAAGACTTGAGGAGAAATGGCTGATCTC 298
Db 337 TGGCAAACTGAGCGCTTAAACACAAATGAAGACTTGAGGAGAAATGGCCGAATCTC 396
QY 299 TCCGAATACCAAGGCCCTATTGATCAGGGGACAGCTACAGAAAGAGTCCGTGTTTATG 358
Db 397 TCCGGATACCAAGGCCCTATTGATCAGGGGCCAGCTATAGGAAGAGTACGGGTTTATG 456
QY 359 AAGAACAGCTTTGTAAGGCCAAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG 418
Db 457 AAGAGCAGCTTTGTAAGGCCAAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG 516
QY 419 ATCTGGAAAGCATCATGAAATCTTAAAGAGGAGATTGAAATGGAGCTAAAGAGCTCT 478
Db 517 GTCTGGGGAAGGATCATGAAATCTCTGAGGAGGAGATTGAAATGGAGCTAAAGAGCTCT 576
QY 479 GGTTTTTTCTACAAAGTGAATTTGAAGAAATTAAGAAATTAAGAAAGAAACGAACTCCAAA 538
Db 577 GGTTTTTTCTACAGAGTGAATTTGAAGAAATTAAGAAATTAAGAAAGAAATGAACCTCCAAA 636
QY 539 GACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAAGGTCTATCATGACAGATC 598
Db 637 GACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAAGGTCTATATATGACGGATC 696
QY 599 TATACCTACCTCAGTCAACAGATGAGCAGGTGAGTGGCGGGAAGAAAGCAAGATC 658
Db 697 TATACCTACCTCAGTCAACAGATGAGCAGGTGAGTGGCGGGAAGAAAGCAAGATC 756
QY 659 TGACAGAGCTGTTCCAGCGGAGATAACATATCTCAGAAATCCCAAGGACTGCGAGCAAG 718
Db 757 TGACAGAACTGTTCCAGCGGAGATAACATATCTTCAGAAATCCCAAGGACTGCGAGCAAG 816
QY 719 CCAGAAAGCTGTTATGTAATCAACAAAGGCTGTGGCTATGGATGTCAATCCATCATG 778
Db 817 CCAGAAAGCTGTTATGTAATCAACAAAGGCTGTGGCTATGGATGTGTGAGTCCATCATG 876
QY 779 TGGTTTACTGCTTCATGATGCTTATGACCCAGCAACACTCATCTTGGAACTCTCAGA 838
Db 877 TGGTTTACTGCTTCATGATGCTTATGACCCAGCAACACTCATCTTGGAACTCTCAGA 936
```

```
QY 839 ATTGGCGCTATGCTACTTGGAGGATGGAGACTGTGTTTAGACCTGTAACTGAGACATGCA 898
Db 937 ATTGGCGCTATGCTACTTGGTGGATGGAGACTGTATTTAGGCCCTGTAACTGAGACATGCA 996
QY 899 CAGACAGTCTGGCCTCTCCACTGGACACTGGTCAGGTGAAGTCAAGGACAAAAATGTTTC 958
Db 997 CAGACAGATCTGGCATCTCCACTGGACACTGGTCAGGTGAAGTCAAGGACAAAAATGTTTC 1056
QY 959 AAGTGGTCGAGCTCCCAATTTGTAGACAGCCTCCATCTCGTCTCTCTTACTTACCTTGG 1018
Db 1057 AAGTGGTCGAGCTTCCCATTTGTAGACAGCTCTTATCCCGTCTCTCCATATTTACCTTGG 1116
QY 1019 CTGTACAGAAAGACCTTGCAGATCGACTCTCTGAGAGTCCATGGTGATCTCTGCAGTGGT 1078
Db 1117 CTGTACAGAAAGACCTTGCAGATCGACTTGTACAGTGCATGTGTGACCTGACCTGCTGGT 1176
QY 1079 GGGTATCCCAAGTTGTCAAATACTTGTATCCGTCCACAACTTGGCTCGGAAAGGAAATAG 1138
Db 1177 GGGTGTCTCAGTTTGTCAAATACTTGTATCCGCCACAGCTTGGCTAGAAAGAAATAG 1236
QY 1139 AAGAAACACCAAGAGCTTGGCTTCAAACATCCAGTTATTGGAGTCCATGTCAGACGCA 1198
Db 1237 AAGAAACACCAAGAGCTTGGCTTCAAACATCCAGTTATTGGAGTCCATGTCAGACGCA 1296
QY 1199 CTGACAAAGTGGGAACAGAAAGCAGCTTCCATCCATTTGAGGAATACATGTCACGTTG 1258
Db 1297 CAGACAAAGTGGGAACAGAAAGCTGCTTCCATCCCATTTGAAGAGTACATGTCGATGTTG 1356
QY 1259 AAGAAACATTTTCAAGTCTCGAACGCGAGAAATGAAGTGGATATAAAGAGAGTGTATCTGG 1318
Db 1357 AAGAAACATTTTCAAGTCTCGAACGCGAGAAATGAAGTGGATATAAAGAGAGTGTATCTGG 1416
QY 1319 CCACTGATGACCCCTTCTTTGTTAAAGGAGGCAAGACAAAGTACTTCCAAATTTATGAATTTA 1378
Db 1417 CCACAGATGACCCCTTCTTTTAAAGGAGGCAAGAAACAAAGTACTTCCAAATTTATGAATTTA 1476
QY 1379 TTAGTGATAACTCTATTTCTTTGTCAGCTGGAATACACACCGGATACACAGAAATTCAC 1438
Db 1477 TTAGTGATAACTCTATTTCTTTGTCAGCTGGAATACACACCGGATACACAGAAATTCAC 1536
QY 1439 TTTGGGGCGGTGATCCTCGATATACACTTCTCTCCAGGCTGACTTCTTGTGTGTACTT 1498
Db 1537 TTTGGGGCGGTGATCCTCGATATACACTTCTCTCTCAGGAGACTTCTTGTGTGTACTT 1596
QY 1499 TTTTCATCCAGGCTCTGTAGGGTTGCTTATGAAATTCATGCAAAACATGCTATCTGATGCT 1558
Db 1597 TTTTCATCCAGGCTCTGTAGGGTTGCTTATGAAATTCATGCAAAACATGCTATCTGATGCT 1656
QY 1559 CTGCAAACTTCCATCTTTTAGATGACATCTATCTATTTTGGAGGCCAAATGCCCAAC 1618
Db 1657 CTGCAAACTTCCATCTTTTAGATGACATCTATCTATTTTGGAGGCCAAATGCCCAAC 1716
QY 1619 AGATTGAGTTTATCTCTCCACCACTCGAACTAAAGAGGAAATCCCATGSAACCTGGAG 1678
Db 1717 AATTTGCAATTTATGCTCACCACCACTCGAACTCGAGATGAATTTCCATGSAACCTGGAG 1776
QY 1679 ATATCATTTGTTGGCTGGAAACCAATTTGGAATGGTTACTCTAAAGGTGTCAACAGAAAC 1738
Db 1777 ATATCATTTGTTGGCTGGAAATCAATTTGGGATGGCTATTTCTAAAGGTGTCAACAGAAAT 1836
QY 1739 TAGGAAACAGGCTGTACCTTCTCTCAAAAGTCCGAGAGAGATAGAAACAGTCAAT 1798
Db 1837 TGGGAAAGGAGGCGCTATATCTCCCTCTCAAAAGTTCGAGAGAGATAGAAACGCTCAAGT 1896
QY 1799 ACCCTACATATCTTGAAGCTGAAAAATAGAGATGGAGTGAAGAGATTAAACACAGAAAT 1858
Db 1897 ACCCTACATATCTTGAAGCTGAAAAATAGAGATGGAGTGAAGAGATTAAACACAGAAAT 1956
QY 1859 TAGTTACAGACCATCTCAG 1876
Db 1957 CAGTTCCAGCAAACTCAG 1974
```

Search completed: February 2, 2004, 16:22:47  
Job time : 7393.83 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 10:17:23 ; Search time 543.385 Seconds  
(without alignments)  
9975.371 Million cell updates/sec

Title: US-09-971-773-1

Perfect score: 2008

Sequence: 1 aacagaaactatttctctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2008	100.0	2008	24	ABK70031 Antibody productio
2	1585.6	79.0	3291	23	ABV22468 Human prostate exp
3	1585.6	79.0	3291	23	ABV28283 Human prostate exp
4	1584	78.9	1728	24	ABK70032 Antibody productio
5	1574.2	78.4	3007	22	AAH33315 Human colon cancer
6	1572.8	78.3	2100	18	AAH76574 Human alpha 1-6 fu
7	1514.8	75.4	1759	24	ABA98809 Alpha1,6-fucosyl t
8	1474.2	73.4	1728	18	AAH76573 Pig alpha 1-6 fuco

9	976.4	48.6	979	21	AAC63891 Chinese hamster FU
10	976.4	48.6	979	24	ABK70036 Antibody productio
11	893.2	44.5	979	21	AAC63892 Rat FUT8 cDNA. Ra
12	893.2	44.5	979	24	ABK70037 Antibody productio
13	881	43.9	1017	22	AAF87952 Human alpha 1-6 fu
14	599.8	29.9	699	22	AAF87953 Human alpha 1-6 fu
15	367.6	18.3	2761	23	ABL04601 Drosophila melanog
16	305.6	15.2	394	20	AAV89049 EST clone CB100.
17	302	15.0	9196	24	ABK70033 Antibody productio
18	285.6	14.2	503	22	AAK11265 Human brain expres
19	285.6	14.2	503	22	AAK11265 used
20	231	11.5	4682	23	ABL04600 Drosophila melanog
21	218.8	10.9	551	22	AAK11037 Human brain expres
22	213.4	10.6	248	22	AAI42659 Probe #11345 used
23	213.4	10.6	248	22	ABA47883 Human breast cell
24	213.4	10.6	248	22	ABA65776 Human foetal liver
25	213.4	10.6	248	22	ABA32861 Probe #11327 for g
26	213.4	10.6	248	22	AAK14177 Human brain expres
27	213.4	10.6	248	22	AAK25142 Human brain expres
28	213.4	10.6	248	22	AAK39911 Human bone marrow
29	213.4	10.6	248	22	AAI20722 Probe #10655 for g
30	213.4	10.6	248	22	AAI45939 Probe #14625 used
31	213.4	10.6	248	22	AAI57183 Probe #25869 used
32	213.4	10.6	248	22	AAI06420 Probe #6411 used t
33	213.4	10.6	248	23	ABS39501 Human liver single
34	213.4	10.6	248	24	ABS14010 Human genome-deriv
35	149	7.4	481	25	ABT21901 Breast cancer mark
36	136.4	6.8	384	22	ABA42755 Human breast cell
37	136.4	6.8	384	22	ABA53183 Human foetal liver
38	136.4	6.8	384	22	ABA22955 Probe #1421 for ge
39	136.4	6.8	384	22	AAK01429 Human brain expres
40	136.4	6.8	384	22	AAK26877 Human bone marrow
41	136.4	6.8	384	22	AAI11508 Probe #1441 for ge
42	136.4	6.8	384	22	AAI32785 Probe #1471 used t
43	136.4	6.8	384	22	AAI01422 Probe #1413 used t
44	136.4	6.8	384	23	ABS26469 Human liver single
45	136.4	6.8	384	24	ABS01477 Human genome-deriv

#### ALIGNMENTS

RESULT 1  
ABK70031  
ID ABK70031 standard; cDNA; 2008 BP.  
XX  
AC ABK70031;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Antibody production method related cDNA #1.  
XX  
KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.  
XX  
OS Cricetulus griseus.  
XX  
FN WO200231140-A1.  
XX  
PD 18-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-JP08804.  
XX  
PR 06-OCT-2000; 2000JP-0308526.  
XX  
PA (KYOWA ) KYOWA HAKKO KOGYO KK.  
XX  
PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX  
XX WPI; 2002-340182/37.



Db 1741 GGAAGAAACAGGCGCTGACCTCTCCCTACAAAGTCCGAGAGAAATAGAAACAGTCAAAATAC 1800  
Qy 1801 CCTACATATCCTGAAGCTGAAAGAAATAGAGATGGAGTGAAGAGATTAACAACAGAAATTA 1860  
Db 1801 CCTACATATCCTGAAGCTGAAAGAAATAGAGATGGAGTGAAGAGATTAACAACAGAAATTA 1860  
Qy 1861 GTTCAGACCATCTCAGCCCAAGCAGAGAACCCAGACCTAACATATGTTTCATTGACAGACAT 1920  
Db 1861 GTTCAGACCATCTCAGCCCAAGCAGAGAACCCAGACCTAACATATGTTTCATTGACAGACAT 1920  
Qy 1921 GCTCGCACCAAGAGCAAGTGGGAACCTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980  
Db 1921 GCTCGCACCAAGAGCAAGTGGGAACCTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980  
Qy 1981 GAAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2008  
Db 1981 GAAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2008

## RESULT 2

ABV22468

ID ABV22468 standard; cDNA; 3291 BP.

XX AC

ABV22468;

XX DT 13-SEP-2002 (first entry)

XX DE

Human prostate expression marker cDNA 22459.

XX KW

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
pharmacogenomic marker; gene; ss.

XX OS

Homo sapiens.

XX PN

WO200160860-A2.

XX PD

23-AUG-2001.

XX PF

20-FEB-2001; 2001WO-US05171.

XX PR

17-FEB-2000; 2000US-183319P.

XX PR

16-MAR-2000; 2000US-189862P.

XX PR

25-MAY-2000; 2000US-207454P.

XX PR

09-JUN-2000; 2000US-211314P.

XX PR

18-JUL-2000; 2000US-219007P.

XX PR

13-DEC-2000; 2000US-255281P.

XX PA

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI

Schlegel R, Endege WO, Monahan JB;

XX PS

WPI; 2001-662795/76.

XX DR

Novel isolated nucleic acid molecule associated with cancerous state of  
prostate cells and correlating with presence of prostate cancer, useful  
for detecting presence of prostate cancer, stage of prostate cancer -

XX PS

Claim 1; Page 3914; 11750pp; English.

XX CC

The invention relates to an isolated nucleic acid molecule (I) comprising  
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
specification or its complement. (I) is useful for:

XX CC

(a) assessing whether a patient is afflicted with prostate cancer;

XX CC

(b) monitoring the progression of prostate cancer in a patient;

XX CC

(c) assessing the efficacy of a test compound to inhibit prostate

XX CC

cancer in a patient;

XX CC

(d) assessing the efficacy of a therapy for inhibiting prostate cancer

XX CC

in a patient;

XX CC

(e) selecting a composition for inhibiting prostate cancer in a patient;

XX CC

(f) assessing the prostate cell carcinogenic potential of a compound;

XX CC

(g) determining whether prostate cancer has metastasized in a patient;

XX CC

(h) assessing the aggressiveness or indolence of prostate cancer in a  
patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX SQ

Sequence 3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other;

Query Match

Best Local Similarity 79.0%; Score 1585.6; DB 23; Length 3291;

Matches 1789; Conservative 0; Mismatches 219; Indels 32; Gaps 4;

Qy 1

AACGAAAGCTTATTTCTGTGTGGCTAACTAGAACACAGAGTACAAATGTTTCCAAATCTT 60  
Db 626 AACGAAAGCTTATTTCACTGTGCACTAAGTAAACAGAGTTACAAATGTTTCCAAATCTT 685

Qy 61

TAGCTCCGAGAAACACA--GAAGGAGTTGAAACTCTGAAATCGGGCATGAGCTGTT 118  
Db 686 TGAGCTCCAGGACCTCCAGGGAAGTGGTTGAAATCTGAAATCGGGCATGAGCTGTT 745

Qy 119

CTGCGGTGGATATGCTCATTTCTTTGCTGGGGACCTTATTTGTTTATATAGTGT 178  
Db 746 CTTGGCGTTGGATTATGCTCATTTCTTTGCTGGGGACCTTCTGTTTATATAGTGT 805

Qy 179

GTCAATTTGTTTCGAGATAATGACCACCTGACCATTTCTAGCAGAGAACTCTCCAAATTC 238  
Db 806 GTCATTTGTTACGAGATAATGACCATCTCTATCTACTAGCCGAACTGTCCAAGATTC 865

Qy 239

TTGCAAGCTGGAGCGCTTAAACACAACAAATGAAGCTTGAGGAGAAATGCTGAGTCTC 298  
Db 866 TGCAAGCTTGAAAGCTTAAACACAGCAGAAATGAAGCTTGAGGCGAATGCGCAATCTC 925

Qy 299

TCCGAATACAGAAAGCCCTTATTCATCAGGGGACAGCTACAGGAAGAGTCTGTTTGTAG 358  
Db 926 TCCGATACCAAGAAAGCCCTTATTCATCAGGGGACAGCTATAGGAAGAGTACGCGTTTGTAG 985

Qy 359

AAGACAGCTTGTAAAGGCCAAAGAACAGATTGAAATTTACAAGAAACAACTAGGAATG 418  
Db 986 AAGAGCAGCTTGTAAAGGCCAAAGAACAGATTGAAATTTACAAGAAACAACTAGGAATG 1045

Qy 419

ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAAATTTGAGCTTAAAGAGCTCT 478  
Db 1046 GTCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAAATTTGAGCTTAAAGAGCTCT 1105

Qy 479

GGTTTTTCTACAAAGTGAATTTGAAGAAATTTAAGAAATTTAAGAGAAACAACTCCAAA 538  
Db 1106 GGTTTTTCTACAGATGAATTTGAAGAAATTTAAGAAATTTAAGAGAAATGAACCTCCAAA 1165

Qy 539

GACATGAGATGAATTTCTTTTGGATTAGACATCATGAAAGTCTCATCATGACAGATC 598  
Db 1166 GACATGAGATGAATTTCTTTTGGATTAGACATCATGAAAGTCTCATCATGACAGATC 1225

Qy 599

TATACCTCAGTCAAAACAGATGGAGCGAGTGGCGGGAAGGAAAGGAAAGGAAAGATC 658  
Db 1226 TATACCTCAGTCAAGACAGATGGAGCGAGTGGCGGGAAGGAAAGGAAAGGAAAGATC 1285

Qy 659

TGACAGAGCTGGTCCAGCGGAGATAACATATCTGAGAAATCCAGAGCTGACGCAAG 718  
Db 1286 TGACAGAGCTGGTCCAGCGGAGATAACATATCTTCAAGATCCCAAGAGCTGACGCAAG 1345

Qy 719

CCAGAAAGCTGGTATGTAATATCAACAAAGCTGTGGCTATGGATGTCACTCCATCATG 778  
Db 1346 CCAAGAAAGCTGGTATGTAATATCAACAAAGCTGTGGCTATGGATGTCACTCCATCATG 1405

Qy 779

TGGTTACTGCTCATCATATCTTATGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 838  
Db 1406 TGGTTACTGCTCATCATATCTTATGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1465

Qy 839

ATTGCGGCTATGCTACTGAGAGGATGGGAGATGTTTGTAGACCTGTAGTGAAGATGCA 898  
Db 1466 ATTGCGGCTATGCTACTGAGAGGATGGGAGATGTTTGTAGGCTGTAGTGAAGATGCA 1525

Qy 899

CAGACAGCTGGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGCAAAATGTTTC 958  
Db 1526 CAGACAGCTGGGCTCTCCACTGACACTGGTCAAGTGAAGTGAAGCAAAATGTTTC 1585

Qy 959

AAGTGGTCGAGCTCCCCATTGTAGACAGCTCCATCTCTCGTCTCTTACTTACCTTTGG 1018



179 GTCAATTTGGTTGAGATAATGACCAACCCCTGACCACTTCTAGCAGAGAACTCTCCAGATTC 238  
1806 GTCACTTGGTACGAGATAATGACCATCTGATCACTCTAGCCGAGAACTGTCCAGATTC 865  
239 TTGCAAGCTGGAGGGCTTAAACCAACAAATGAAGCTTGGAGGAGATGGCTGAGTCTC 298  
866 TGGCAAGCTTTGAACGCTTAAACAGCAGAAATGAAGACTTTGAGGCGAATGGCCGAATCTC 925  
299 TCCGAATACCAAGAGGCCCTTATGATCAGGGGACAGCTTACAGGAAGAGTCCGCTGTTTAG 358  
926 TCCGATACCAAGAGGCCCTTATGATCAGGGGACAGCTTACAGGAAGAGTCCGCTGTTTAG 985  
359 AAGAACAGCTTTGTTAAGSCCAAGAACAGATTTGAAATTTACAGAAACAAAGCTAGGAATG 418  
986 AAGAGCAGCTTTGTTAAGSCCAAGAACAGATTTGAAATTTACAGAAACAAAGCTAGGAATG 1045  
419 ATCTGGGAAAGGATCATGAATCTTTAAGGAGGAGATTTGAAATTTGAGAGCTTAAAGAGCTCT 478  
1046 GTCTGGGAAAGGATCATGAATCTTTAAGGAGGAGATTTGAAATTTGAGAGCTTAAAGAGCTCT 1105  
479 GGTCTTTTCTACAAAGTGAATTTGAAGAAATTTAAGAAATTTAAGAGGAAACGAACTCCAAA 538  
1106 GGTCTTTTCTACAGATGATTTGAAGAAATTTAAGAAATTTAAGAGGAAATGAACTCCAAA 1165  
539 GACATGCAGATGAAATCTCTTTTGGATTTAGGACATCATGAAAGGTCTATCATGACAGATC 598  
1166 GACATGCAGATGAAATCTCTTTTGGATTTAGGACATCATGAAAGGTCTATATGACGGATC 1225  
599 TATCTACTCTCAGTCAAAACAGATGAGCAGGTGAGTGGCGGGGAAAGAAAGACCCAAAGATC 658  
1226 TATCTACTCTCAGTCAAAACAGATGAGCAGGTGATTTGGCGGGGAAAGAAAGACCCAAAGATC 1285  
659 TGACAGAGCTGGTCCAGCGGAGATAACATATCTGCAGAAATCCCAAGGACTGACGCAAG 718  
1286 TGACAGAACTGGTTCAGCGGAGATAACATATCTTCAGAAATCCCAAGGACTGACGCAAG 1345  
719 CCAGAAAGCTGGTATGTAATTAACAAAGGCTGTGGCTATGGATGTCAACTCCATCATG 778  
1346 CCAGAAAGCTGGTATGTAATTAACAAAGGCTGTGGCTATGGCTGTGCTCCATCATG 1405  
779 TGGTTTACTGTTCTCATGATGCTTATGGCACCCAGCGAAACCTCATCTTGGAAATCTCAGA 838  
1406 TGGTCTACTGTTCTCATGATGCTTATGGCACCCAGCGAAACCTCATCTTGGAAATCTCAGA 1465  
839 ATTTGGGCTATGCTTACTGAGAGTGGAGACTGTGTTTAGACCTGTAGTACGACATGCA 898  
1466 ATTTGGGCTATGCTTACTGAGAGTGGAGACTGTATTTTAGGCGCTGTAAAGTACGACATGCA 1525  
899 CAGACAGGCTTGGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAAATGTTT 958  
1526 CAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGTGAAGTGAAGGACAAAATGTTT 1585  
959 AAGTGGTTCGAGCTCCCATGTTAGACAGCTCCATCTCTGCTCTCTCTTACTTACCTTTGG 1018  
1586 AAGTGGTTCGAGCTTCCCATGTTAGACAGCTTCTATCCCGCTCTCCATATTTACCTTTGG 1645  
1019 CTGTACCAAGAGACCTTCGAGATCGACTCTCTGAGAGTCCATGGTGCATCTCCAGTGGT 1078  
1646 CTGTACCAAGAGACCTTCGAGATCGACTGTGTACGAGTGCATGGTGACCTTCAGTGGT 1705  
1079 GGGTATCCAGTTCCTCAAAATCTTGAATCCGTCACAACTTGGCTGGAAAGGAAATAG 1138  
1706 GGGTCTCTCAGTTTGTCAATATCTTGTATCCGCCACAGCTTGGCTAGAAAAGAAATAG 1765  
1139 AAGAAACCAACCAAGAGCTTGGCTTCAAAACATCCAGTTATTGGAGTCCATGTCCAGCGCA 1198  
1766 AAGAGGCCCAACCAAGAGCTTGGCTTCAAAACATCCAGTTATTGGAGTCCATGTCCAGCGCA 1825  
1199 CTGACAAAGTGGGAAACAGAGCAGGCTTCCATCCCATTTAGGAAATACATGTCACGTTG 1258  
1826 CAGACAAAGTGGGAAACAGAGGCTTCCATCCCATTTGAAAGAGTACATGGTGCATGTTG 1885  
1259 AAGAACATTTTTCAGCTTCTCGAAACGACAGATGAAAGTGAATAAAAAAGAGTGTATCTGG 1318

## RESULT 4

ABK70032

ID ABK70032 standard; cDNA; 1728 BP.

XX AC ABK70032;

XX DT 15-JUL-2002 (first entry)

XX XX Antibody production method related cDNA #2.

XX KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
XX KW antibacterial; antinflammatory; anti allergic; allergy; inflammation;  
XX KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
XX OS tumour; circulatory disease; infection; primer; ss.

XX OS Mus musculus.

XX XX WO200231140-A1.

XX XX 18-APR-2002.

XX XX

1886 AAGAACATTTTTCAGCTTCTTTCGACGCAAGATGCAAGTGGCAAAAAAGAGTGTATTTGG 1945  
1319 CCAGTATGACCCCTCTCTTCTTAAAGGAGGCAAGACAAAGTACTTCCAAATATCAATTTTA 1378  
1946 CCAGATGACCCCTCTCTTATTAAGGAGGCAAAACAAAGTACTCCCAATATCAATTTTA 2005  
1379 TTAGTGATAAATCTCTATTTCTTGGTTCAGCTGGACTTACAAACCGGATACACAGAAATTCAC 1438  
2006 TTAGTGATAAATCTCTATTTCTTGGTTCAGCTGGACTTACAAATCGATACACAGAAATTCAC 2065  
1439 TTGGGGGCGTGATCTCTGGATATACATTTCTCTCCAGGCTGACTTCTTGTGTGTACTTT 1498  
2066 TTCTGTGAGTGATCTCTGGATATACATTTCTCTCTCAGGCGAGCTTCTTAGTGTGTACTTT 2125  
1499 TTTTCATCCAGGCTGTAGGTTGCTTATGAAATCATGCAAAACACTGCTCATCTCTGATGCTT 1558  
2126 TTTTCATCCAGGCTGTAGGTTGCTTATGAAATCATGCAAAACACTGCTCATCTCTGATGCTT 2185  
1559 CTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGAGGCCAAAATGCCCAACACC 1618  
2186 CTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGGGGCCAGAAATGCCCAACACC 2245  
1619 AGATTGCAAGTTTATCTCTCAACCACTGAACTTAAAGAGGAAATCCCAATGGAACCTGGAG 1678  
2246 AAATTGGCAATTTATGTTTCAACCACTGCAAGTGAATTTCCCATGGAACCTGGAG 2305  
1679 ATATCATTTGTTGGTGGAAATCCATTTGGAATGTTTACTCTTAAAGGTGTCAACAGAAAC 1738  
2306 ATATCATTTGTTGGTGGAAATCCATTTGGAATGTTTACTCTTAAAGGTGTCAACAGAAAT 2365  
1739 TAGGAAAAACAGGCGCTGATCCCTTCTTCAAAAGTCCGAGAGAGATAGAAACAGTCAAAAT 1798  
2366 TGGGAAGGACGGGCTTATATCCCTCTTCAAAAGTTCGAGAGAGATAGAAACGCTCAAGT 2425  
1799 ACCCTACATATCTCTGAAAGCTGAAAAATAGAGATGGAGTGTAAAGATTAA----- 1848  
2426 ACCCAATATCTCTGAGGCTGAGAAATAAAGCTCAGATGGAAGAGATAGAAACAGCAAACT 2485  
1849 -----CAACAGATTTAGTTTCAGACCATCTCAGCCCAAGCAGAACCCAG----- 1893  
2486 CAGTTCCACCAAACTCAGTTTCAACCAATTTTCCAGCAAACTGTAGATGAAGAGGCTCTGA 2545  
1894 ----ACTAACATATGTTTCATTTGACAGACATGCTCCGACACCAAGAGCAAGTGGAAACCT 1949  
2546 TCTAACAAATAAGTTTATATGATAGTACTCTCTCAGACCAACAGCAGCTGGGAACTGA 2605  
1950 CAGATGCTGC-ACCTGTGGAAACGCTCTTTGTGAAGGGCTGCTGTGCCCTCAAGCCCATG 2008  
2606 CATAGGCTTCAATTTGGTGGAAATCTCTTTAAACAAGGGCTGCAATGCCCTCATACCCCATG 2665

```
PF 05-OCT-2001; 2001WO-JP08804.
XX
PR 06-OCT-2000; 2000JP-0308526.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
PA
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
XX WPI; 2002-340182/37.
XX
PT Cells producing antibody compositions including antibody fragments and
PT fusion proteins with Fc domain of antibody, useful for prevention or
PT treatment of cancer, immune diseases, circulatory diseases and
PT infections
XX
XX Claim 32; Page 4-11; 314pp; Japanese.
XX
CC This invention relates to novel method for antibody production
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell
CC transfected with a gene encoding an antibody molecule for producing a
CC composition comprising an antibody molecule with an Fc domain bonded
CC to the N-glycoside linkage complex sugar chain. The produced antibody
CC compositions are drugs for prevention or treatment of diseases
CC accompanying tumour, allergy or inflammation, autoimmune diseases,
CC circulatory diseases, and viral and bacterial infections. The
CC antibodies can be stably produced using the method of the invention
CC with high binding activity and potency thus leading to high safety and
CC reduced side effects when applied alone or in combination with other
CC drugs for therapy. The present sequence represents a nucleotide
CC molecule used in the method of the invention.
XX
SQ Sequence 1728 BP; 529 A; 365 C; 400 G; 434 T; 0 other;

Query Match 78.9%; Score 1584; DB 24; Length 1728;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 100 ATGCGGGATGACCTGGTTCCTGGGCTGGGATGATGCTCAATCTTTTGGCTGGGGGACC 159
DB 1 ATGCGGGGATGACCTGGTTCCTGGGCTGGGATGATGCTCAATCTTTTGGCTGGGGGACC 60

QY 160 TTATTTGTTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCCCTGACCATCTTAGC 219
DB 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCCCTGATCACTCCAGC 120

QY 220 AGAGAACTCTCAAGATCTTTCGAAGCTGGAGCGCTTAAACACAAATGAAGACTTG 279
DB 121 AGAGAACTCTCAAGATCTTTCGAAGCTTGAACGCTTAAACACAAATGAAGACTTG 180

QY 280 AGGAGAACTGGCTGAGTCTCTCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACA 339
DB 181 AGGCGAATGGCTGAGTCTCTCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACA 240

QY 340 GGAAGAGCTCCGTTGTTTTAGAGAACAGCTTGTAAAGGCCAAAGAACAGATGAAATTAAC 399
DB 241 GGAAGAGCTCCGTTGTTTTAGAGAACAGCTTGTAAAGGCCAAAGAACAGATGAAATTAAC 300

QY 400 AAGAACAAGCTAGGAATGATCTGGGAAGGATCATGAATCTTAAGAGGAGGATTTGAA 459
DB 301 AAGAACAAGCTAGGAATGATCTGGGAAGGATCATGAATCTTAAGAGGAGGATTTGAA 360

QY 460 AATGAGCTTAAAGAGCTCTGGTTTCTTACAAAGTGAATTAAGAAATTAAGAAATTA 519
DB 361 AATGAGCTTAAAGAGCTCTGGTTTCTTACAAAGTGAATTAAGAAATTAAGAAATTA 420

QY 520 GAAGGAAACGAACCTCAAAGACATCCAGATGAATCTTTGGATTTAGGACATCATGAA 579
DB 421 GAAGGAAATGAACCTCAAAGACATCCAGATGAATCTTTGGATTTAGGACATCATGAA 480

QY 580 AGGCTTATCATGACAGATCTATACCTCAGTCAACAGATGGAGCAGGTCAGTGGCGG 639
DB 481 AGGCTTATCATGACAGATCTATACCTCAGTCAACAGATGGAGCAGGTCAGTGGCGG 540
```

```
640 GAAAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCGAGAA 699
641 GAAAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCGAGAA 600
700 CCCAAGGACTGCAGCAAGGCCAGAAAGCTGTATGTAAATATCAACAAAGGCTGTGGCTAT 759
601 CCTAAGGACTGCAGCAAGGCCAGAAAGCTGGTGTAAACATCAATTAAGGCTGTGGCTAT 660
760 GGATGTCAACTCCATCATGTGTTTACTGCTTCAATGATTGCTTATGGCACCACCGAACA 819
661 GGTGTCAACTCCATCATGTGTTTACTGCTTCAATGATTGCTTATGGCACCACCGAACA 720
820 CTCAATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGAGGATGGGAGACTGTGTTAGA 879
721 CTCAATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGCTGATGGGAGACTGTGTTAGA 780
880 CCTGTAAGTGAGACATGCACAGACAGCTCTGGCCCTCTCCACTGGACACTGGTCAGGTGAA 939
781 CCTGTAAGTGAGACATGCACAGACAGATCTGGCCCTCTCCACTGGACACTGGTCAGGTGAA 840
940 GTGAAGGACAAATAATGTTCAAGTGGTCTGAGCTCCCAATTTGTAGACAGCTCCATCTCGT 999
841 GTAAATGACAAATAATGTTCAAGTGGTCTGAGCTCCCAATTTGTAGACAGCTCCATCTCGG 900
1000 CCTCTTACTTACCCCTTGGCTGTACCAAGAGACCTTGCAGATCGACTCTCGAGAGTCCAT 1059
901 CCTCTTACTTACCCCTTGGCTGTTCAGAAAGACCTTGCAGACCCGACTCTTAAGAGTCCAT 960
1060 GGTGATCTCTGAGTGGTGGTATCCAGTTGCTCAATCTGCTCAATCTGATCGGTCACACACT 1119
961 GGTGACCTCTGAGTGGTGGTGTCCAGTTGCTCAATCTGCTCAATCTGATTCGTCACACACT 1020
1120 TGGCTGGAAGGGAAATAGAAAGAAACCAAGAGAGCTTGGCTTCAACATCTCCAGTTAT 1179
1021 TGGCTGGAAGGGAAATAGAAAGAAACCAAGAGAGCTTGGCTTCAACATCTCCAGTTAT 1080
1180 GGAGTCCATGTACAGACCTGACAAAGTGGGAAACAGAAAGCAGCTTCCACCCCATCGAG 1239
1081 GGAGTCCATGTACAGACCTGACAAAGTGGGAAACAGAAAGCAGCTTCCACCCCATCGAG 1140
1240 GAATACATGTTACAGCTTGAAGACATTTTCAGCTTCTCGAACGACAGAAAGTGGAT 1299
1141 GAGTACATGTTACAGCTTGAAGACATTTTCAGCTTCTCGAACGACAGAAAGTGGAT 1200
1300 AAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTTAAAGAGGCAAGACAAAG 1359
1201 AAAAAAGAGTGTATCTGGCTACTGATGATCTTCTTTGTTTAAAGAGGCAAGACAAAG 1260
1360 TACTCCAAATTAAGATTTATAGTGATACTCTATTTCTTGGTCAGCTGGACTACAAAC 1419
1261 TACTCCAAATTAAGATTTATAGTGATACTCTATTTCTTGGTCAGCTGGACTACAAAC 1320
1420 CGATACAGAAATTTCACTTCGGGGGCTGATCTCGGATATACACTTCTCTCCAGGCT 1479
1321 CGGTACAGAAATTTCACTTCGGGGGCTGATCTCGGATATACACTTCTCTCCAGGCT 1380
1480 GACTTCTTGTGTACTTTTCACTCCAGCTCTGTAGGGTGTGCTTATGAATCATGCAA 1539
1381 GACTTCTTGTGTACTTTTCACTCCAGCTCTGTAGGGTGTGCTTATGAATCATGCAA 1440
1540 AACTGCTATCTGATGCTCTGAAACCTTCAATCTTTTATAGATGATCTACTATTTTGA 1599
1441 ACCCTGCTATCTGATGCTCTGAAACCTTCAATCTTTTATAGATGATCTACTATTTTGA 1500
1600 GGCCAAATGGCCCAACACAGATTTGATCTCCACCAACCTCCAGCTTAAGAGGAA 1659
1501 GGCCAAATGGCCCAACACAGATTTGATCTCCACCAACCTCCAGCTTAAGAGGAA 1560
1660 ATCCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCAATGGGAATGGTTACTCT 1719
1561 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCAATGGGAATGGTTACTCT 1620
```









277	Qy	GTCTCTGGTACGAGATAATGACCATCTCTGATCATCTCTAGCCGAGAACTGTCCAAAGATTC	336
239	Qy	TTGCAAAAGCTGGAGCGCTTAAAAACAAACAAATGAAGACTTTGAGGAGAAATGGCTCAGTCTC	298
337	Db	TTGCAAAAGCTTGAACGCTTAAACACAGCAGAAATGAAGACTTTGAGGCGAAATGGCCGAATCTC	396
299	Qy	TCCGAATACAGAAAGGCCCTTATTGATCAGGGGACAGCTACAGGAAGANGTCCGTGTTTTAG	358
397	Db	TCCGGATACCAGAAAGGCCCTTATTGATCAGGGGCCAGCTATAGGAAGAGTAGCGGTTTTAG	456
359	Qy	AAGAAACAGCTTGTTTAAGGCCAAAGAACAGAAATGAAAAATTACAAAGAACAAAGCTAGGAATG	418
457	Db	AAGAGCAGCTTGTTAAGGCCAAAGAACAGAAATGAAAAATTACAAAGAACAAAGCAGCAAAATG	516
419	Qy	ATCTCGGAAAGGATCATGAAATCTTAAGGAGGAGGATTTGAAAAATTGGAAGCTATAAGAGCTCT	478
517	Db	GTCTGGGGAAGGATCATGAAATCTGAGGAGGAGGATTTGAAAAATTGGAAGCTATAAGAGCTCT	576
479	Qy	GGTTTTTTTACAAAAGTGAAATTGAAGAAATTTAAGAAATTTAGAAAGGAAACGAATCCAAA	538
577	Db	GGTTTTTTTCTACAGAGTGAATTTGAAGAAATTTAAGAAATTTAGAAAGGAAATGAAATCCAAA	636
539	Qy	GACATGCAGATGAAATTTCTTTTGGATTTAGGACATCATGAAAGGTCTATCATCAGCAGATC	598
637	Db	GACATGCAGATGAAATTTCTTTGGATTTAGGACATCATGAAAGGTCTATTAATGACCGATC	696
599	Qy	TATACTACCTCAGTCAAAACAGATGGACGAGTGAAGTGGCGGGAAGAAAGAACGCAAGATC	658
697	Db	TATACTACCTCAGTCAAGACAGATGGACGAGTGAAGTGGCGGGAAGAAAGAACGCAAGATC	756
659	Qy	TGACAGAGCTGGTCCAGCGGAGAAATAATATCTGCAGAAATCCCAAGGACTGCAGCAAAAG	718
757	Db	TGACAGAACTGGTTTCAGCGGAGAAATAATATCTTCAGAAATCCCAAGGACTGCAGCAAAAG	816
719	Qy	CCAGAAAGCTGGTATGTAATAATCAACAAAGCTGTGGCTATGGATGCACTCCATCATG	778
817	Db	CCAAAAAGCTGGTGTGTAATAATCAACAAAGCTGTGGCTATGGCTGTGCACTCCATCATG	876
779	Qy	TGGTTTACTGTCTCATGATTCCTTATGGCACCCAGCGAACACTCATCTTCGAAATCTCAGA	838
877	Db	TGGTCTACTGCTTCATGATTCGATATGGCACCCAGCGAACACTCATCTTCGAAATCTCAGA	936
839	Qy	ATTGGCGCTATGCTACTCGGAGATGGGAGACTGTGTTTAGACCTGTGAAGTGAGACATGCA	898
937	Db	ATTGGCGCTATGCTACTGGTGGATGGGAGACTGTATTTAGCGCTGTGAAGTGAGACATGCA	996
899	Qy	CAGACAGGTCTGGCCTCTCCACTCGACACACTGCTCAGGTGAAGTGAAGGACAAAAAATGTTT	958
997	Db	CAGACAGATCTGGCATCTCCACTCGACACACTGCTCAGGTGAAGTGAAGGACAAAAAATGTTT	1056
959	Qy	AAGTGGTGCAGCTCCCAATGTAGACAGCTCCATCTCGTCCCTCTTACTTACCCCTTGG	1018
1057	Db	AAGTGGTGCAGCTTCCCATTTAGACAGCTCTTCAATCCCGTCTCTCCATATTTACCCCTTGG	1116
1019	Qy	CTGTACCAAGAGACTTTCGAGATCGACTCTCTGAGAGTCCATGGTGTACTCTCGAGTGTGGT	1078
1117	Db	CTGTACCAAGAGACTTCGAGATCGACTTGTGACGAGTGCATGGTGAACCTCTCGAGTGTGGT	1176
1079	Qy	GGGTATCCCGATTTGTCAAAATACCTTGATTCGGTCCACAACTTGTGCTGGAAAGGGAAATAG	1138
1177	Db	GGGTGTCTCAGTTTTGTCAAAATACCTTGATTCGGTCCACAACTTGTGCTGGAAAGGGAAATAG	1236
1139	Qy	AAGAAACACCAAGAGCTTGGCTTCAACATCCAGTTATTGGAGTCCATGTCAGACGCA	1198
1237	Db	AAGAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATTGGAGTCCATGTCAGACGCA	1296
1199	Qy	CTGACAAAGTGGGAACGAAGCAGCTTCCATCCCATTTAGGAAATACATGGTACACGCTTG	1258
1297	Db	CAGACAAAGTGGGAACGAAGCAGCTTCCATCCCATTTAGGAGTACATGGTGCATGTTG	1356
1259	Qy	AAGAACATTTTTCAGCTTCTCGAAACGCAAGATGAAAGTGGATATAAAAAGAGTGTATCTGG	1318
1357	Db	AAGAACAATTTTTCAGCTTCTTTCAGCGCAAGATGCAAGTGGACAAAAAAGAGTGTATTTGG	1416

QY	1319	CCACTGATGACCCCTCTCTTGTAAAGAGGCGAAAGACAAAGTACTCTCAATTAATGAATTTA	1379
DB	1417	CCACAGATGACCCCTCTCTTATTTAAAGAGGCGAAAGACAAAGTATACCCCAATTAATGAATTTA	1476
QY	1379	TTAGTGATAACTCTATTTCTTTGGTGCAGCTGGACTACACAACCGGATACACAGAAAAATTCAC	1438
DB	1477	TTAGTGATAACTCTATTTCTTTGGTGCAGCTGGACTGGACATCGACATCGATACACAGAAAAATTCAC	1536
QY	1439	TTGGGGCGTGATCCCTCGATATACATTTTCTCTCCAGGTGACTTCCCTTGTGTGTAATCTT	1498
DB	1537	TTCTGTGAGTGATCCCTCGATATACATTTTCTCTCAGGCGAGACTTCTCTAGTGTGTACTT	1596
QY	1499	TTTTCATCCAGGTCTGTAGGGTTGCTTTATGAAATCATGCAAACTGCATCGATCCTGATGCCT	1558
DB	1597	TTTTCATCCAGGTCTGTAGGGTTGCTTTATGAAATCATGCAAACTGCATCGATCCTGATGCCT	1656
QY	1559	CTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGAGGCGCAAAATGCCCAACAAC	1618
DB	1657	CTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGGGGGCCAGATGCCCAACAAC	1716
QY	1619	AGATTGAGTTTATCTCTCAACCACTCGAACTAAAGAGGAAATCCCCATGGAACCTTGAG	1678
DB	1717	AAATTGCCATTTATGCTTCAACCAACCCGGAACCTGCAGATGAAATTTCCCATGGAACCTTGAG	1776
QY	1679	ATATCATTTGGTGTGGTGGAAACCATTTGGAATGGTACTCTTAAAGGTGTCAACAGAAAC	1738
DB	1777	ATATCATTTGGTGTGGTGGAAACCATTTGGAATGGTACTCTTAAAGGTGTCAACAGAAAC	1836
QY	1739	TAGGAAAAACAGGCGCTGTACCCCTTCTTACAAAAGTCCGAGAGAGATAGAAAACTGCAAAAT	1798
DB	1837	TGGAGGAGCGGGCCATATCCCTCTTACAAAAGTTCGAGAGAGATAGAAAACTGCAAAAT	1896
QY	1799	ACCTCATATCTCTGAGCTGAAAAATAGAGATGGAGTGTAAAGATTTAACACAGAAAT	1858
DB	1897	ACCCACATATCTCTGAGCTGAGAAATAAAGCTCAGATGGAAGAGATAAACGACCAAACT	1956
QY	1859	TAGTTTCAGACCATCTCTAG 1876	
DB	1957	CAGTTTCAGCAAACTCTAG 1974	
RESULT 7			
ID	ABA98809	standard; cDNA; 1759 BP.	
XX	AC	ABA98809;	
XX	AC	ABA98809;	
DT	07-MAY-2002	(first entry)	
XX	DE	Alphal,6-fucosyl transferase cDNA.	
XX	XX	Plant; glycoprotein; alphal,6-fucosyl transferase; alphal,6-Ft;	
KW	KW	enzyme; ss.	
XX	OS	Homo sapiens.	
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	17..1744	
FT	FT	/*tag= a	
FT	FT	/product= "Alphal,6-fucosyl transferase"	
XX	XX		
FN	JP2001333787-A.		
XX	XX		
PD	04-DEC-2001.		
XX	XX		
PF	06-MAR-2001; 2001JP-0062704.		
XX	XX		
PR	22-MAR-2000; 2000JP-0081059.		
XX	XX		
PA	(TANI//) TANIGUCHI N.		
PA	(SEKI//) SEKI T.		
PA	(FUJI//) FUJIYAMA K.		



XX 05-MAR-1998 (first entry)  
 XX Pig alpha 1-6 fucosyltransferase gene.  
 XX Alpha 1-6 fucosyltransferase; enzyme; pig; human; glucose transfer;  
 KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GlcNAc; cancer diagnosis; ss.  
 XX Sus scrofa.  
 XX Key Location/Qualifiers  
 FT 1..1728  
 FT CDS /\*tag= a  
 XX WO9727303-A1.  
 XX 31-JUL-1997.  
 XX 23-JAN-1997; 97WO-JP00171.  
 XX 22-JUL-1996; 96JP-0192260.  
 PR 24-JAN-1996; 96JP-0010365.  
 PR 21-JUN-1996; 96JP-0161648.  
 PR 24-JUN-1996; 96JP-0162813.  
 XX (TOYM) TOYO BOSEKI KK.  
 XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
 PI 1997-393690/36.  
 DR P-PSDB; AAW22124.  
 XX Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 PT synthesis and modification of sugar chains and used as an antigen  
 PT for production of diagnostic antibodies  
 XX Claim 5; Page 30-34; 61pp; Japanese.  
 CC AAT76573 and AAT76574 represent the coding sequences for the pig and  
 CC human alpha 1-6 fucosyltransferases of the invention, respectively. The  
 CC encoded enzyme transfers fucose from guanosine diphosphate to the  
 CC 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:  
 CC (GlcNAc-beta 1-2Man-alpha 1-6)(GlcNAc-beta 1-2Man-alpha 1-3)Man-beta  
 CC 1-4GlcNAc-beta 1-4GlcNAc-R to give (GlcNAc-beta 1-2Man-alpha 1-6)(GlcNAc-beta  
 CC 1-2Man-alpha 1-3)Man-beta 1-4GlcNAc-beta 1-4(Fu-alpha 1-6)GlcNAc-R. It has  
 CC an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the  
 CC pH range 4-10 after 5 hours at 4 degrees C. The optimum working  
 CC temperature of the enzyme is 30-37 degrees C. A bivalent metal is not  
 CC required for activity of the enzyme, and the enzyme is not inhibited in  
 CC the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 CC modification of sugar chains, and as antigen for the production of  
 CC antibodies recognising the enzyme. The antibodies can be used for the  
 CC diagnosis of cancer and other diseases.  
 XX Sequence 1728 BP; 521 A; 362 C; 419 G; 426 T; 0 other;  
 SQ  
 Query Match 73.4%; Score 1474.2; DB 18; Length 1728;  
 Best Local Similarity 90.9%; Pred. No. 0;  
 Matches 1569; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
 QY 100 ATGCGGGCATGGACGTTCCTGGGTTGGATTATGCTCATCTCTTTTGGCTGGGGACC 159  
 DB 1 ATGCGGCCCATGGACGTGGTTCGGGCTGGATTATGCTCATCTCTTTTGGCTGGGGACC 60  
 QY 160 TTATTGTTTTATATAGGTGGTTCATTTGGTTCCAGATAATGACACCCCTGACCATCTAGC 219  
 DB 61 TTGCTATTTTACATAGTGGTTCATTTGGTACGAGATATGACCATCTGTATCACTCTAGC 120  
 QY 220 AGAGAACTCTCCAGATTCTTCCAAAGCTGGAGCGCTTTAAACAAACAAATGAAGACTTG 279  
 DB 121 CGAGAACTGCTCCAGATTCTTGGCAAAGCTGGAAACGCTTTAAACAAACAAATGAAGACTTG 180

QY 280 AGGAAATGGCTAGTCTCTCCGAATACCAGAGGCGCTATTGTATCAGGGGACAGCTACA 339  
 DB |||||  
 DB 181 AGGAAATGGCTGAATCTCTCCGAATACCAGAGGCGCTATTGTATCAGGGGCGAGCTTCA 240  
 QY 340 GGAAGAGTCCGTGTTTTAGAGAACAGCTTGTTHAAGGCCAAAGAACAGATTGAAATTTAC 399  
 DB |||||  
 DB 241 GGAAGAGTCCGTGTTTTAGAGAACAGCTTGTTHAAGGCCAAAGAACAGATTGAAATTTAT 300  
 QY 400 AAGAAACAAAGCTAAGAAATGATCTGGAAAGAGTATCATGAATATCTTTAAGGAGGAGATTGAA 459  
 DB |||||  
 DB 301 AAGAAACAAAGCTAAGAAATGATCTGGAAAGAGTATCATGAATATCTTTAAGGAGGAGATTGAA 360  
 QY 460 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTA 519  
 DB |||||  
 DB 361 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTA 420  
 QY 520 GAAGAAACAAAGCTAAGAAATGATCTGGAAAGAGTATCATGAATATCTTTAAGGAGGAGATTGAA 579  
 DB |||||  
 DB 421 GAAGAAACAAAGCTAAGAAATGATCTGGAAAGAGTATCATGAATATCTTTAAGGAGGAGATTGAA 480  
 QY 580 AGGTCTATCATGACAGATCTATCTACTCTCAGTCAAACAGATGGAGCAGGTGAGTGGCGG 639  
 DB |||||  
 DB 481 AGGTCTATCATGACAGATCTATCTACTCTCAGTCAAACAGATGGAGCAGGTGAGTGGCGT 540  
 QY 640 GAAAGAAAGCCAAAGATCTGACAGAGTGGTCCAGCGGAGAAATAACATATCTTCAGAAAT 699  
 DB |||||  
 DB 541 GAAAGAGGAGCCAAAGATCTGACAGAGTGGTCCAGCGGAGAAATAACATATCTTCAGAAAT 600  
 QY 700 CCCAAGGAGTGCAGTAAAGCCAGAAAGCTGTATGTATATATCAACAAAGGCTGTGGCTAT 759  
 DB |||||  
 DB 601 CCCAAGGAGTGCAGTAAAGCCAGAAAGCTGTATGTATATCAACAAAGGCTGTGGCTAT 660  
 QY 760 GGATGTCAACTCCATCATGTGGTTTACTGCTTCTCATGTTTCTTATGGCACCAGCGGAACA 819  
 DB |||||  
 DB 661 GGCTGTGAGTCCCATCATGTAGTGTACTGCTTTTATGATTTGATATGGCACCAGCGGAACA 720  
 QY 820 CTCAATCTTGAATCTCAGAAATGGCGCTATGCTACTTGGAGATGGGAGACTGTGTTTAGA 879  
 DB |||||  
 DB 721 CTGCGCTTGGAACTCTCACAATTTGGCGCTACGCTACTGGGGATGGGAACTGTGTTTAGA 780  
 QY 880 CCTGTAGTGTAGACATGACAGACAGAGTCTGGCTCTCCACTGGACACACTGTGAGTGAA 939  
 DB |||||  
 DB 781 CCTGTAGTGTAGACATGACAGACAGAGTCTGGCAGCTCCACTGGACACACTGTGAGTGAA 840  
 QY 940 GTGAGGACACAAATGTTCAAGTGTGAGCTGCTCCCAATGTTAGACAGCTCCATCTCGT 999  
 DB |||||  
 DB 841 GTAAAGGACACAAATGTTCAAGTGTGAGCTGCTCCCAATGTTAGACAGCTTCATCTCGT 900  
 QY 1000 CCTCTTACTTTACCCCTTGGCTGTATCCAGAAAGCTTGCAGATCGACTCCTGAGAGTCCAT 1059  
 DB |||||  
 DB 901 CCTCCATATTTACCCCTTGGCTGTATCCAGAAAGCTTGCAGATCGACTTGTACGAGTCCAT 960  
 QY 1060 GGTGATCTCGAGTGTGGTGGTATCCAGTTTGTCAAATATCTTGTATCCGTCACAACT 1119  
 DB |||||  
 DB 961 GGTGATCTCGAGTGTGGTGGTATCCAGTTTGTCAAATATCTTGTCAAGTACTTGTTCGCCACAACT 1020  
 QY 1120 TGGCTGGAAGGGAATAGAGAAACCAACCAAGAGCTTGGCTTCAACATCCAGTTAT 1179  
 DB |||||  
 DB 1021 TGGCTGGAAGGGAATAGAGAAAGGAGGCCCAAGAGCTTAGGCTTCAACATCCAGTTAT 1080  
 QY 1180 GGAGTCCATGTGACAGCTGACACACAGAGTGGGAAACAGAGCAGCTTCCATCCCATTCAG 1239  
 DB |||||  
 DB 1081 GGAGTCCATGTGACAGCTGACACACAGAGTGGGAAAGGAGCAGCTTCCATCCCATTCAG 1140  
 QY 1240 GAATACATGTTACAGTGTGAAGAAACATTTTCAGCTTCTCGAAGCAGAGTGAAGTGGAT 1299  
 DB |||||  
 DB 1141 GAATACATGTTACAGTGTGAAGAAAGCTTTCAGCTTCTTCGTCAGAAATGCAAGTGGAT 1200  
 QY 1300 AAAAAAGAGTGTATCTGGCCACTGTACCCCTTCTTTGTTTAAAGGAGGCAAGCAAG 1359  
 DB |||||  
 DB 1201 AAAAAAGAGTGTATCTGGCCACTGTACCCCTTCTTTGTTTAAAGGAGGCAAGCAAG 1260  
 QY 1360 TACTCCAATTATGAATTTATTTAGTGAATACTCTATTCTTGTGTCAGCTGGACTACACAAC 1419

Db 1261 TACCCAGTATGAATTTATAGTATGATATCTATCTCTTGTCAGCTGACATATAT 1320  
Qy 1420 CGATACACAGAAATTCACCTCGGGCGTATCCTGGATATACACTTCTCTCCAGGCT 1479  
Db 1321 CGATATACAGAAATTCACCTCGGGGTGATCCTGGATATACACTTCTCTCCAGGCA 1380  
Qy 1480 GACTTCTTGTTGTTACTTTTATCCAGGCTCTGAGGTTGCTTATGAATCATGCAA 1539  
Db 1381 GACTTCTTGTTGTTACTTTTATCCAGGCTCTGAGGTTGCTTATGAATCATGCAA 1440  
Qy 1540 ACATGTCATCTGATCCCTCCAACTTCCATCTTTAGATGACATCTACTATTTTGA 1599  
Db 1441 GCGCTGATCCCTGATGCTCTCGGAATTCCTTTGGATGACATCTACTATTTTGA 1500  
Qy 1600 GGCCTAAATGCCCAACAGATTCAGTTATCTCTCAACCTCGAATTAAGAGGAA 1659  
Db 1501 GGCCTAAATGCCCAACAGATTCAGTTATCTCTCAACCTCGAATTAAGAGGAA 1560  
Qy 1660 ATCCCATGGAACCTGGAGATATCATTTGGTGGCTGGAAACCATTTGAAATGGTTACTCT 1719  
Db 1561 ATCCCATGGAACCTGGAGATATTTTGGTGGCTGGAAATCATCTGGGATGGCTATCTCT 1620  
Qy 1720 AAGGTGTCAACAGAAACTAGAAACAGGCTGTACCTTCTTACAAAGTCCGAGAG 1779  
Db 1621 AAGGTGTCAACAGAAACTAGAAACAGGCTGTACCTTCTTACAAAGTCCGAGAG 1680  
Qy 1780 AAGATAGAAACAGTCAATACCTTACATATCTGAAAGTGAATAAATA 1826  
Db 1681 AAGATAGAAACAGTCAATACCTTACATATCTGAAAGTGAATAAATA 1727

## RESULT 9

AAC63891  
ID AAC63891 standard; cDNA; 979 BP.  
XX AAC63891;  
XX  
DT 09-FEB-2001 (first entry)  
XX Chinese hamster FUT8 cDNA.  
DE  
XX Immunologically functional molecule; immune system; immunomodulation;  
KW glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;  
KW viral infection; bacterial infection; allergy; autoimmune disease;  
KW inflammation; antibody; Chinese hamster; FUT8; ss.  
XX  
OS Cricetulus griseus.  
XX  
PN WO200061739-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 07-APR-2000; 2000WO-JP02260.  
XX  
PR 09-APR-1999; 95JP-0103158.  
XX  
PA (KYOW) KYOWA HAKKO KOGYO KK.  
XX  
PI Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;  
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;  
XX  
DR WPI; 2000-665129/64.  
XX  
PT Control of activity of antibodies and other immunologically functional  
PT molecules by addition or removal of fucose from sugar chain for  
PT diagnosis and treatment of cancer, allergy and other diseases -  
XX  
PS Example 8; Page 72-73; 81pp; Japanese.  
XX  
CC The invention relates to a method for controlling the activity of an  
CC immunologically functional molecule (e.g., an antibody) where the  
CC control is effected by the presence or absence of fucose bound to an

CC N-acetylglucosamine residue at the reducing end of the sugar chain on  
CC the immunologically functional molecule. The invention also relates to  
CC methods for the diagnosis, prevention or treatment of diseases which  
CC involve the modified immunologically functional molecule, and agents  
CC which stimulate the activity of an immunologically functional molecule.  
CC The methods of the invention are used for the diagnosis, treatment and  
CC prevention of a broad range of diseases including cancer, circulatory  
CC disease, viral or bacterial infection, allergy, autoimmune disease and  
CC inflammation. The present sequence represents a Chinese hamster FUT8  
CC cDNA isolated in an exemplification of the invention.  
XX  
SQ Sequence 979 BP; 286 A; 227 C; 218 G; 248 T; 0 other;

Query Match 48.6%; Score 976.4; DB 21; Length 979;  
Best Local Similarity 99.9%; Pred. No. 7.6e-280;  
Matches 977; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 819 ACTCATCTTGAATCTCAGAAATTCGCGCTATGCTACTGGAGGATGGGAGACTGTGTTAG 878  
Db 1 ACTCATCTTGAATCTCAGAAATTCGCGCTATGCTACTGGAGGATGGGAGACTGTGTTAG 60  
Qy 879 ACCTGTAAGTGAGACATGCACAGACAGGTCTGGGCTCTCCACTGGACACTGGTCAGGTGA 938  
Db 61 ACCTGTAAGTGAGACATGCACAGACAGGTCTGGGCTCTCCACTGGACACTGGTCAGGTGA 120  
Qy 939 AGTGAAGACAAAATGTTCAAGTGGTGGAGTCCCATTTGTAGACAGCTCCATCTCTCG 998  
Db 121 AGTGAAGACAAAATGTTCAAGTGGTGGAGTCCCATTTGTAGACAGCTCCATCTCTCG 180  
Qy 999 TCCTCCTTACTTACCTTGGCTGTACAGAAAGACTTGCAGATCGACTCTCTGAGAGTCCA 1058  
Db 181 TCCTCCTTACTTACCTTGGCTGTACAGAAAGACTTGCAGATCGACTCTCTGAGAGTCCA 240  
Qy 1059 TGGTGATCCTGCAGTGTGGGTATCCAGTTTGTCAAAATPACTTGATCCGTCACAAACC 1118  
Db 241 TGGTGATCCTGCAGTGTGGGTATCCAGTTTGTCAAAATPACTTGATCCGTCACAAACC 300  
Qy 1119 TTGGCTGAAAGGGAATAGAAAGAACCAAGAGCTGGGCTTCAAAATCCAGTTAT 1178  
Db 301 TTGGCTGAAAGGGAATAGAAAGAACCAAGAGCTGGGCTTCAAAATCCAGTTAT 360  
Qy 1179 TGGAGTCCATGTACAGCGCTGACAAAGTGGGAACAGAGCAGCTTCCATCCATTTGA 1238  
Db 361 TGGAGTCCATGTACAGCGCTGACAAAGTGGGAACAGAGCAGCTTCCATCCATTTGA 420  
Qy 1239 GGAATACATGTTACACGTTGAAGAACATTTTCAAGTCTCTCGAACGCGAATGAAGTGA 1298  
Db 421 GGAATACATGTTACACGTTGAAGAACATTTTCAAGTCTCTCGAACGCGAATGAAGTGA 480  
Qy 1299 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGAGGCAAGACAAA 1358  
Db 481 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGAGGCAAGACAAA 540  
Qy 1359 GTACTCCAATTTATGAATTTATAGTATTAATCTCTTTTGGTTCAGTGCAGTACACAA 1418  
Db 541 GTACTCCAATTTATGAATTTATAGTATTAATCTCTTTTGGTTCAGTGCAGTACACAA 600  
Qy 1419 CCGATACACAGAAATTCACCTTCGGGCGTGTATCTCGATATACACTTCTCTCCAGGC 1478  
Db 601 CCGATACACAGAAATTCACCTTCGGGCGTGTATCTCGATATACACTTCTCTCCAGGC 660  
Qy 1479 TGACTTCTTGTTGTTACTTTTTCATCCAGGCTGTAGGGTTCCTTATGAATCATGCA 1538  
Db 661 TGACTTCTTGTTGTTACTTTTTCATCCAGGCTGTAGGGTTCCTTATGAATCATGCA 720  
Qy 1539 AACACTCATCTGATGCTCTCAAACTTCCATTTTGTAGATGACATCTACTATTTTGG 1598  
Db 721 AACACTCATCTGATGCTCTCAAACTTCCATTTTGTAGATGACATCTACTATTTTGG 780  
Qy 1599 AGGCCAAAATGCCCAACAGATTTGAGTTTATCTCACCACCTCGAACTAAAGAGGA 1658  
Db 781 AGGCCAAAATGCCCAACAGATTTGAGTTTATCTCACCACCTCGAACTAAAGAGGA 840

QY 1659 AATCCCATGAACCTGGAGATATCATTTGGTGTGGCTGGAACCACTTGGAAATGGTTACTC 1718  
DB |||||  
841 AATCCCATGAACCTGGAGATATCATTTGGTGTGGCTGGAACCACTTGGAAATGGTTACTC 900  
QY 1719 TAAAGGTGTCAACAGAAAACCTAGGAAAACAGGCTGTACCTTCTTCAAAAGTCCGAGA 1778  
DB |||||  
901 TAAAGGTGTCAACAGAAAACCTAGGAAAACAGGCTGTACCTTCTTCAAAAGTCCGAGA 960  
QY 1779 GAAGATAGAAACAGTCAA 1796  
DB |||||  
961 GAAGATAGAAACGGTCAA 978

## RESULT 10

ABK70036

ID ABK70036 standard; cDNA; 979 BP.

XX

AC

XX

XX

15-JUL-2002 (first entry)

XX Antibody production method related cDNA #4.

DE

XX

KW

KW

KW

KW

KW

XX

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
tumour; circulatory disease; infection; primer; ss.

Cricetulus griseus.

WO200231140-A1.

18-APR-2002.

05-OCT-2001; 2001WO-JP08804.

06-OCT-2000; 2000JP-0308526.

(KYOWA) KYOWA HAKKO KOGYO KK.

Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;

Hosaka E, Yamano K, Yamasaki M, Hanai N;

WPI; 2002-340182/37.

Cells producing antibody compositions including antibody fragments and  
fusion proteins with Fc domain of antibody, useful for prevention or  
treatment of cancer, immune diseases, circulatory diseases and  
infections

Example 9; Page 12-13; 314pp; Japanese.

This invention relates to novel method for antibody production  
comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
transferred with a gene encoding an antibody molecule for producing a  
composition comprising an antibody molecule with an Fc domain bonded  
to the N-glycoside linkage complex sugar chain. The produced antibody  
compositions are drugs for prevention or treatment of diseases  
accompanying tumour, allergy or inflammation, autoimmune diseases,  
circulatory diseases, and viral and bacterial infections. The  
antibodies can be stably produced using the method of the invention  
with high binding activity and potency thus leading to high safety and  
reduced side effects when applied alone or in combination with other  
drugs for therapy. The present sequence represents a nucleotide  
molecule used in the method of the invention.

Sequence 979 BP; 286 A; 227 C; 218 G; 248 T; 0 other;

Query Match 48.6%; Score 976.4; DB 24; Length 979;

Best Local Similarity 99.9%; Pred. No. 7.6e-280;

Matches 977; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 819 ACTCATCTTGGAAATCTCAGAAATGGCGTATGCTACTGGAGGATGGAGACTGTGTTAG 878

DB |||||  
QY 1 ACTCATCTTGGAAATCTCAGAAATGGCGTATGCTACTGGAGGATGGAGACTGTGTTAG 60  
DB |||||  
QY 879 ACCTGTAAGTGAGACATGACACAGAGTCTGGCTCTCCACATGGACACTGGTTCAGGTGA 938  
DB |||||  
QY 61 ACCTGTAAGTGAGACATGACACAGAGTCTGGCTCTCCACATGGACACTGGTTCAGGTGA 120  
DB |||||  
QY 939 AGTGAAGCAGCAAAAAGTTCAAGTGGTTCGAGCTCCCAATTTGTAGACAGCTCCATCCTCG 998  
DB |||||  
QY 121 AGTGAAGCAGCAAAAAGTTCAAGTGGTTCGAGCTCCCAATTTGTAGACAGCTCCATCCTCG 180  
DB |||||  
QY 999 TCCTCCTTACTTACCTTTGGCTGTACAGAGAGCTTTGCAGATCGACTCTCTGAGAGTCCA 1058  
DB |||||  
QY 181 TCCTCCTTACTTACCTTTGGCTGTACAGAGAGCTTTGCAGATCGACTCTCTGAGAGTCCA 240  
DB |||||  
QY 1059 TGGTGATCCTGCAGTGTGGTGGTATCCAGTTTGTCAAAATACTTGTATCCGTCCACAACC 1118  
DB |||||  
QY 241 TGGTGATCCTGCAGTGTGGTGGTATCCAGTTTGTCAAAATACTTGTATCCGTCCACAACC 300  
DB |||||  
QY 1119 TTGGCTGGAAGGGAATAGAAACACCAAGAGCTTGGCTTCAAAATCCAGTTAT 1178  
DB |||||  
QY 301 TTGGCTGGAAGGGAATAGAAACACCAAGAGCTTGGCTTCAAAATCCAGTTAT 360  
DB |||||  
QY 1179 TGGAGTCCATGTGAGACGCACTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATTTGA 1238  
DB |||||  
QY 361 TGGAGTCCATGTGAGACGCACTGACAAAGTGGGAAACAGAGCAGCTTCCATCCCATTTGA 420  
DB |||||  
QY 1239 GGAATACATGGTACACCTTGAAGAACATTTTTCAGCTTCTCGAACGCGAAGTGAAGTGA 1298  
DB |||||  
QY 421 GGAATACATGGTACACCTTGAAGAACATTTTTCAGCTTCTCGAACGCGAAGTGAAGTGA 480  
DB |||||  
QY 1299 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTTAAAGAGGCAAGACAAA 1358  
DB |||||  
QY 481 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTTAAAGAGGCAAGACAAA 540  
DB |||||  
QY 1359 GTACTCCAATTTATGAATTTTATAGTGAATCTTATTTTGTGGTTCAGCTGACATACAAA 1418  
DB |||||  
QY 541 GTACTCCAATTTATGAATTTTATAGTGAATCTTATTTTCTGGTTCAGCTGACATACAAA 600  
DB |||||  
QY 1419 CCGATACACAGAAAATTCACCTTCGGGGGTGATCTCGATATACACTTCTCTCCAGGC 1478  
DB |||||  
QY 601 CCGATACACAGAAAATTCACCTTCGGGGGTGATCTCGATATACACTTCTCTCCAGGC 660  
DB |||||  
QY 1479 TGACTTCTCTGTGTACTTTTTCATCCAGGTCTGTAGGGTTCCTTATGAAATCATGCA 1538  
DB |||||  
QY 661 TGACTTCTCTGTGTACTTTTTCATCCAGGTCTGTAGGGTTCCTTATGAAATCATGCA 720  
DB |||||  
QY 1539 AACACTGCATCCTGATCTCTGCAAACTTCCATTTTATAGATGACATCTACTATTTTGG 1598  
DB |||||  
QY 721 AACACTGCATCCTGATCTCTGCAAACTTCCATTTTATAGATGACATCTACTATTTTGG 780  
DB |||||  
QY 1599 AGGCCAAAATGCCCAACACAGATTCGAGTTTATCTCCACCAACTCGAACTAAAGAGA 1658  
DB |||||  
QY 781 AGGCCAAAATGCCCAACACAGATTCGAGTTTATCTCCACCAACTCGAACTAAAGAGA 840  
DB |||||  
QY 1659 AATCCCATGAACTCGAGATATCATTTGGTGTGGCTGGAACCACTTGGAAATGGTTACTC 1718  
DB |||||  
QY 841 AATCCCATGAACTCGAGATATCATTTGGTGTGGCTGGAACCACTTGGAAATGGTTACTC 900  
DB |||||  
QY 1719 TAAAGGTGTCAACAGAAAACCTAGGAAAACAGGCTGTACCTTCTTCAAAAGTCCGAGA 1778  
DB |||||  
QY 901 TAAAGGTGTCAACAGAAAACCTAGGAAAACAGGCTGTACCTTCTTCAAAAGTCCGAGA 960  
DB |||||  
QY 1779 GAAGATAGAAACAGTCAA 1796  
DB |||||  
QY 961 GAAGATAGAAACGGTCAA 978

## RESULT 11

AAC63892

ID AAC63892 standard; cDNA; 979 BP.

XX

AC

```

XX 09-FEB-2001 (first entry)
XX Rat FUT8 cDNA.
XX Immunologically functional molecule; immune system; immunomodulation;
KW glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;
KW viral infection; bacterial infection; allergy; autoimmune disease;
KW inflammation; antibody; rat; FUT8; ss.
XX Rattus sp.
XX WO200061739-A1.
XX 19-OCT-2000.
XX 07-APR-2000; 2000WO-JP02260.
XX 09-APR-1999; 99JP-0103158.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;
XX WPI; 2000-665129/64.
XX Control of activity of antibodies and other immunologically functional
PT molecules by addition or removal of fucose from sugar chain for
PT diagnosis and treatment of cancer, allergy and other diseases -
XX Example 8; Page 73-74; 81pp; Japanese.
XX The invention relates to a method for controlling the activity of an
CC immunologically functional molecule (e.g., an antibody) where the
CC control is effected by the presence or absence of fucose bound to an
CC N-acetylglucosamine residue at the reducing end of the sugar chain on
CC the immunologically functional molecule. The invention also relates to
CC methods for the diagnosis, prevention or treatment of diseases which
CC involve the modified immunologically functional molecule, and agents
CC which stimulate the activity of an immunologically functional molecule.
CC The methods of the invention are used for the diagnosis, treatment and
CC prevention of a broad range of diseases including cancer, circulatory
CC disease, viral or bacterial infection, allergy, autoimmune disease and
CC inflammation. The present sequence represents a rat FUT8 cDNA isolated
CC in an exemplification of the invention.
XX Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;
SQ
Query Match 44.5%; Score 893.2; DB 21; Length 979;
Best Local Similarity 94.6%; Pred. No. 4.5e-255;
Matches 925; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 819 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGAGGATGGGAGACTGTTTAG 878
DB 1 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGAGGATGGGAGACTGTTTAG 60
QY 879 ACCTCTAAGTAGACATGCACAGACAGCTGTGGCTCTCCACTGCACACTGGTCAGGTGA 938
DB 61 ACCTCTAAGTAGACATGCACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA 120
QY 939 AGTGAAGGACAAATAATGTTCAAGTGGTGGAGCTCCCTCCATTTGTAGACAGCTCCATCTCG 998
DB 121 AGTGAATGACAAATAATTCAGTGGTGGAGCTCCCTCCATTTGTAGACAGCTTCATCTCG 180
QY 999 TCCTCCTTACTTACCGTTGGCTGTACACAGACACTTCGAGATCGACTTCGTGAGGTCCA 1058
DB 181 GCCTCCTTACTTACCAGTGGTGTTCACAGAGACCTTCGAGATCGACTCGTAAGAGTCCA 240
QY 1059 TGGTGATCTCGAGTGGTGGGTATCCAGTTTGTCAAAATCTTGATCCCGTCCACAACC 1118
DB 241 TGGTGATCTCGAGTGGTGGGTGTCCAGTTCGTCAAAATTTGATTCGTCCACAACC 300

```

```

QY 1119 TTGGCTGGAAGGAAATAGAGAAACCAACCAAGAGCTTGGCTTCAACATCCAGTTAT 1178
DB 301 TTGGCTGGAAGGAAATAGAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 360
QY 1179 TGGAGTCCATGTCAGACGCACTGACAAAGTGGGAAACAGACGAGCTTCCATCCATTGA 1238
DB 361 TGGAGTCCATGTCAGACGCACTGACAAAGTGGGAAACAGACGAGCTTCCATCCATTGA 420
QY 1239 GGAATACATGTCACACGTTGAAGAACATTTTTCAGCTTCTCGAACGCAAGATGAAGTGA 1298
DB 421 AGAGTACATGTCACATGTTGAAGAACATTTTTCAGCTTCTCGAACGCAAGATGAAGTGA 480
QY 1299 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTTAAAGAGGCAAGACAAA 1358
DB 481 TAAAAAAGAGTGTATCTGGCTACCGATGACCCCTTCTTTGTTAAAGAGGCAAGACAAA 540
QY 1359 GTACTCCAAATTTATGAATTTATTAGTGATACTCTATTCTTGGTCAGCTGGACTACACAA 1418
DB 541 GTACTCCAAATTTATGAATTTATTAGTGATACTCTATTCTTGGTCAGCTGGACTACACAA 600
QY 1419 CCGATACACAGAAATTCACCTTCGGGGCGTGATCCTGGATATACACTTTCTCCCAGGC 1478
DB 601 TCGGTACACAGAAATTCACCTTCGGGGCGTGATCCTGGATATACACTTTCTCTCAGGC 660
QY 1479 TGACTTCTTGTGTGTACTTTTTCATCCAGGCTGTAGGTTGCTTATGAATCATGCA 1538
DB 661 TGACTTCTTGTGTGTACTTTTTCATCCAGGCTGTGCGGGTTCGTTATGAATCATGCA 720
QY 1539 AACACTCCTCCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 1598
DB 721 AACCTGATCCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 780
QY 1599 AGGCCAAAATGCCCAACACAGATTGCGAGTTTATCTCACCACCTCGAACTAAAGAGGA 1658
DB 781 AGGCCAAAATGCCCAACACAGATTGCGAGTTTATCTCACCACCTCGAACTGATGAGGA 840
QY 1659 AATCCCATGGAACCTGGAGATATCATTTGTTGGCTGGAAACCATTTGGAATGGTTACTC 1718
DB 841 AATTTCAATGGAACCTGGAGATATCATTTGTTGGCTGGAAACCATTTGGGATGGTTATTC 900
QY 1719 TAAAGGTGTCAACAGAAACCTTAGGAAAAACAGGCTGTACCCCTTCTTACAAAGTCCGAGA 1778
DB 901 TAAAGGTGTCAACAGAAACCTTAGGAAAAACAGGCTGTATATCCCTCCTACAAAGTCCGAGA 960
QY 1779 GAAGATAGAAACAGTCAA 1796
DB 961 GAAGATAGAAACGCTCAA 978

```

RESULT 12  
ABK70037  
ID ABK70037 standard; cDNA; 979 BP.

AC ABK70037;

XX 15-JUL-2002 (first entry)

XX Antibody production method related cDNA #5.

XX Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.

OS Rattus norvegicus.

XX WO200231140-A1.

PN 18-APR-2002.

XX 05-OCT-2001; 2001WO-JP08804.

XX 06-OCT-2000; 2000JP-0308526.

```
XX (KYOM ) KYOMA HAKKO KOGYO KK.
XX
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
XX WPI; 2002-340182/37.
XX
XX Cells producing antibody compositions including antibody fragments and
XX fusion proteins with Fc domain of antibody, useful for prevention or
XX treatment of cancer, immune diseases, circulatory diseases and
XX infections
XX
XX Example 9; Page 13-14; 314pp; Japanese.
XX
XX This invention relates to novel method for antibody production
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell
XX transfected with a gene encoding an antibody molecule for producing a
XX composition comprising an antibody molecule with an Fc domain bonded
XX to the N-glycoside linkage complex sugar chain. The produced antibody
XX compositions are drugs for prevention or treatment of diseases
XX accompanying tumour, allergy or inflammation, autoimmune diseases,
XX circulatory diseases, and viral and bacterial infections. The
XX antibodies can be stably produced using the method of the invention
XX with high binding activity and potency thus leading to high safety and
XX reduced side effects when applied alone or in combination with other
XX drugs for therapy. The present sequence represents a nucleotide
XX molecule used in the method of the invention.
XX
XX Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;
XX
XX
XX Query Match 44.5%; Score 893.2; DB 24; Length 979;
XX Best Local Similarity 94.6%; Pred. No. 4.5e-255;
XX Matches 925; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
XX
XX 819 ACTCATCTTGAATCTCAGAATGGCGTATGCTACTGAGAGTGGGAGACTGTTTGTAG 878
XX 1 ACTCATCTTGAATCTCAGAATGGCGTATGCTACTGAGAGTGGGAGACTGTTTGTAG 60
XX
XX 879 ACCTGTAAGTGAGACATGACAGACAGAGTGGCGCTCTCCACTGACACTGGTCAAGTGA 938
XX
XX 61 ACCTGTAAGTGAGACATGACAGACAGAGTGGCGCTCTCCACTGACACTGGTCAAGTGA 120
XX
XX 939 AGTGAAGACAAATATGTTCAAGTGGTGGAGTCCCGCTTGTAGACGCTCCACTCTCG 998
XX 121 AGTGAATGACAAATATGTTCAAGTGGTGGAGTCCCGCTTGTAGACGCTTCATCTCG 180
XX
XX 999 TCCTCTCTTACTTACCTTGGCTGTACCAAGAACCTTGCAGATCGACTCTCGAGAGTCCA 1058
XX 181 GCCTCTTACTTACCACTGGCTGTTCCAGNAGACCTTGCAGATCGACTCGTAAGAGTCCA 240
XX
XX 1059 TGGTGATCTCGAGTGGTGGGATGCCAGTTTGTCAAAATCTTGTATCGGTCACAAACC 1118
XX 241 TGGTGATCTCGAGTGGTGGGATGCCAGTTTGTCAAAATCTTGTATCGGTCACAAACC 300
XX
XX 1119 TTGGCTGGAAGGGAATAGAGAAACCAACAGAGCTTGGCTTCAACATCCAGTTAT 1178
XX 301 TTGGCTAGAAAAGGAAATAGAGAGAGCCCAACAGAGCTTGGCTTCAACATCCAGTCAT 360
XX
XX 1179 TGGAGTCCATGTGACAGCGCACTGACAAAGTGGGAACAGAGCAGGCTTCCATCCCATTTGA 1238
XX 361 TGGAGTCCATGTGACAGCGCACTGACAAAGTGGGAACAGAGCAGGCTTCCATCCCATTTGA 420
XX
XX 1239 GGAATACATGTGTACAGTTGAAGAAATTTTCAGTTTCTCGAAACGCAAGTGAAGTGA 1298
XX 421 AGAGTACATGTGTACAGTTGAAGAAATTTTCAGTTTCTCGAAACGCAAGTGA 480
XX
XX 1299 TAAAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTTAAAGGAGGCAAGACAAA 1358
XX 481 TAAAAAAGAGTGTATCTGGCTACCGATGACCCCTTCTTTTAAAGGAGGCAAGACAAA 540
XX
XX 1359 GTACTCCCAATTATGATTTATTAGTGATAACTCTATTTCTTGGTCAGCTGGACTACAAA 1418
XX
```

```
Db 541 GTACTCCCAATTATGAATTTATTAGTAGATAAATCTTAATTTCTTGGTGCAGCTGCAGTACACAA 600
Qy 1419 CCAGTACACAGAAATTCACCTTCGGGGCGTGATCCTCGATATACACTTTCTCTCCAGGC 1478
Db 601 TCGGTACACAGAAATTCACCTTCGGGGCGTGATCCTCGATATACACTTTCTCTCAGGC 660
Qy 1479 TGACTTCTTGTGTGTACTTTTTCATCCAGGTCTGTAGGGTGTCTTATGAAATCATGCA 1538
Db 661 TGACTTCTTGTGTGTACTTTTTCATCCAGGTCTGTAGGGTGTCTTATGAAATCATGCA 720
Qy 1539 AACACTGCATCTGATGCCCTCTGCAAACTTCCATCTTTAGATGACATCTACTATTTGG 1598
Db 721 AACCTGCATCTGATGCCCTCTGCAAACTTCCATCTTTAGATGACATCTACTATTTGG 780
Qy 1599 AGGCAAAATGCCCAACACAGATTGCGATTATCTCCACCACTCGAACTAAAGAGA 1658
Db 781 AGGCAAAATGCCCAACACAGATTGCGATTATCTCCACCACTCGAACTGATGAGA 840
Qy 1659 AATCCCATGGAACCTCGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGTTTCT 1718
Db 841 AATCCCATGGAACCTCGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGTTTCT 900
Qy 1719 TAAAGGTGTCAACAGAAACTAGGAAAACAGGCTGTACCTTCTTACAAAGTCCGAGA 1778
Db 901 TAAAGGTGTCAACAGAAACTAGGAAAACAGGCTGTATCCCTCTTACAAAGTCCGAGA 960
Qy 1779 GAAGATAGAAACAGTCAA 1796
Db 961 GAAGATAGAAACAGTCAA 978
XX
XX RESULT 13
XX AAF87952
XX ID AAF87952 standard; DNA; 1017 BP.
XX
XX AAF87952;
XX AC AC
XX DT 20-JUL-2001 (first entry)
XX
XX Human alpha 1-6 fucosyltransferase 237-575 encoding DNA SEQ ID NO:1.
XX
XX Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
XX alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;
XX ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1017
XX /*tag= a
XX /partial
XX /product= "alpha 1-6 fucosyltransferase 237-575"
XX
XX JP2001011097-A.
XX
XX 16-JAN-2001.
XX
XX 29-JUN-1999; 99JP-0183569.
XX
XX 29-JUN-1999; 99JP-0183569.
XX (FURE ) FUJIREBIO KK.
XX
XX WPI; 2001-275926/29.
XX P-PSDB; AAB75061.
XX
XX Novel anti-human alpha1-6 fucose transferase antibody useful for
XX immunoassay -
XX
XX Example 1; Page 6-7; lipp; Japanese.
XX
XX The present invention describes an anti-human alpha 1-6
XX fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose
XX
```



CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunoassay for detecting human  
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunoassay. The anti-human  
CC alpha 1-6 fucT antibody can be used for immunoassay. The present  
CC sequence encodes the human alpha 1-6 fucT protein of residues 237 to  
CC 575, which is used in an example from the present invention.  
XX  
SQ Sequence 1017 BP; 296 A; 231 C; 228 G; 262 T; 0 other;

Query Match 43.9%; Score 881; DB 22; Length 1017;  
Best Local Similarity 91.6%; Pred. No. 28-251; Indels 0; Gaps 0;  
Matches 932; Conservative 0; Mismatches 85;

QY 808 ACCGAGCAACTCATCTTGAATCTCAGAAATGGCGTACTGCTACTGGAGATGGAG 867  
DB 1 ACCGAGCAACTCATCTTGAATCTCAGAAATGGCGTACTGCTACTGGATGGAG 60

QY 868 ACTGTTTGTAGACCTGTAAGTGAGACATGCAAGACAGTCTGGCTCTCCACTGGACAC 927  
DB 61 ACTGTTTGTAGACCTGTAAGTGAGACATGCAAGACAGTCTGGCTCTCCACTGGACAC 120

QY 928 TGTGAGGTGAGTGAAGACAAAATGTTCAAGTGGTGGAGTCCCATTTGTAGACAGC 987  
DB 121 TGTGAGGTGAGTGAAGACAAAATGTTCAAGTGGTGGAGTCCCATTTGTAGACAGT 180

QY 988 CTCCATCTCTGCTCTCTTACTTACCTTGGCTGTACCAAGACAGTCTGGACATGCATC 1047  
DB 181 CTTCATCCCGTCTCTTACTTATCCCTTGGCTGTACCAAGACAGTCTGGACATGCATC 240

QY 1048 CTGAGAGTCCATGTTGATCTCTGAGTGGTGGTATCCAGTTTGTCAAAATCTTGATC 1107  
DB 241 GTACGAGTGCATGGTGACCTGAGTGGTGGTGTCTCAGTTTGTCAAAATCTTGATC 300

QY 1108 CGTCCACACCTTGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAA 1167  
DB 301 CGCCACACAGCTTGGCTGGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAA 360

QY 1168 CATCCAGTTATTTGGAGTGCATCTCAGACGCACTGACAAAGTGGGAACAGAGCCCTTC 1227  
DB 361 CATCCAGTTATTTGGAGTGCATCTCAGACGCACTGACAAAGTGGGAACAGAGCCCTTC 420

QY 1228 CATCCATTTGAGGAATACATGTTGACAGTGAAGAACTTTTTCAGCTCTCGAAGCAGA 1287  
DB 421 CATCCATTTGAGGAATACATGTTGACAGTGAAGAACTTTTTCAGCTCTCGAAGCAGA 480

QY 1288 ATGAAAGTGGATAAAAGAGTGTATCTGGCAGTGTGACCTCTTTTGTAAAGGAG 1347  
DB 481 ATGAAAGTGGATAAAAGAGTGTATCTGGCAGTGTGACCTCTTTTGTAAAGGAG 540

QY 1348 GCAAGACAAAGTACTCCAAATATGAATTTATTAGTGATACTCTATTTCTTGGTCAGCT 1407  
DB 541 GCAAGACAAAGTACTCCCAATATGAATTTATTAGTGATACTCTATTTCTTGGTCAGCT 600

QY 1408 GGACTACAAACCGATACACAGAAATTAACCTCGGGCGGTATCCTCGATATACATTT 1467  
DB 601 GGACTACAAACCGATACACAGAAATTAACCTCGGGCGGTATCCTCGATATACATTT 660

QY 1468 CTCTCCAGGTGATCTCTTGTGTACTTTTTCATCCAGGTCTGTAGGGTTCCTAT 1527  
DB 661 CTCTCTCAGGCGAGTCTCTTGTGTACTTTTTCATCCAGGTCTGTAGGGTTCCTAT 720

QY 1528 GAAATCATGCAAACTGATCTCTGATGCTCTGCAAACTTCCATTTTATAGATGACATC 1587  
DB 721 GAAATCATGCAAACTGATCTCTGATGCTCTGCAAACTTCCATTTTATAGATGACATC 780

QY 1588 TACTATTTTGGAGGCAAAATGCCCAACAGATTTGCAAGTTATTCCTCAACCACTCGA 1647  
DB 781 TACTATTTTGGAGGCAAAATGCCCAACAGATTTGCAAGTTATTCCTCAACCACTCGA 840

QY 1648 ACTAAGAGGAATCCCATGGAACCTCGAGATATCATTTGTGTGGCTGGAAACCATGG 1707  
DB 841 ACTGAGATGAATTTCCCATGGAACCTCGAGATATCATTTGTGTGGCTGGAAACCATGG 900

QY 1708 AATGGTTACTTAAGGTGTCAACAGAAAACACTAGAAAACAGGCCTGTACCTTCTTAC 1767  
DB 901 GATGGCTATTCTTAAGGTGTCAACAGAAAATTTGGAAGGACGGCCTATATATCCCTCTTAC 960

QY 1768 AAGTCCGAGAGATAGAAAACAGTCAATACCTACATATCTCTGAAGCTGAAAAA 1824  
DB 961 AAGTTCGAGAGATAGAAAACGGTCAAGTACCCACATATCTCTGAGGCTGAGAAA 1017

RESULT 14  
AAF87953  
ID AAF87953 standard; DNA; 699 BP.  
XX  
AC AAF87953;  
XX  
DT 20-JUL-2001 (first entry)  
XX  
DE Human alpha 1-6 fucosyltransferase 343-575 encoding DNA SEQ ID NO:2.  
XX  
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;  
KW ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..699  
FT /\*tag= a  
FT /partial  
FT /product= "alpha 1-6 fucosyltransferase 343-575"  
XX  
PN JP2001011097-A.  
XX  
PD 16-JAN-2001.  
XX  
PF 29-JUN-1999; 99JP-0183569.  
XX  
PR 29-JUN-1999; 99JP-0183569.  
XX  
PA (FURE ) FUJIREBIO KK.  
XX  
DR WPI; 2001-275926/29.  
DR P-PSDB; AAB/5062.  
XX  
PT Novel anti-human alpha1-6 fucose transferase antibody useful for  
XX immunoassay -  
PS Example 1; Page 7-8; 1lpp; Japanese.  
XX  
CC The present invention describes an anti-human alpha 1-6  
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose  
CC transferase) antibody. Also described are: (1) a hybridoma producing  
CC the above monoclonal antibody; (2) an immunoassay for detecting human  
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;  
CC and (3) a reagent used for the above immunoassay. The anti-human  
CC alpha 1-6 fucT antibody can be used for immunoassay. The present  
CC sequence encodes the human alpha 1-6 fucT protein of residues 343 to  
CC 575, which is used in an example from the present invention.  
XX  
SQ Sequence 699 BP; 220 A; 152 C; 147 G; 180 T; 0 other;

Query Match 29.9%; Score 599.8; DB 22; Length 699;  
Best Local Similarity 91.1%; Pred. No. 8.8e-168;  
Matches 637; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1126 GAAAGGAATAGAGAAACCAACCAAGAGCTTGGCTTCAACATCCAGTTATGGAGTC 1185  
DB 1 GAAAGGAATAGAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATGGAGTC 60

QY 1186 CATGTCCAGCGCACTGACAAAGTGGGAACAGAGAGCTTCCATCCCATTTGAGGAATAC 1245  
DB 61 CATGTCCAGCGCACTGACAAAGTGGGAACAGAGAGCTTCCATCCCATTTGAGGAATAC 120





Db 1794 GCACGTGTGGCTCGCCGGATTTTCCTCGCATCGGATGATGCCAGGTTATGAGGAGCGC 1853  
QY 1352 AGACAAAGTACTCCCAATTATGAATTTATTAGTGATAACTCTATTTCTTTGGTCAGCTGGAC 1411  
Db 1854 GCCGAAGTACCCGAGTACACAGATCATTGATCCGGAGGTGGCGCATGGCGTCGG 1913  
QY 1412 TACACAACCGATACACAGAAAATTCACCTCGGGGGGTGATCCTGGATATACACTTTTCTCT 1471  
Db 1914 TATCCACACGATACACGACACCGCCCTGAACGGGATCATTCTGGATATCCACCTGCTTT 1973  
QY 1472 CCCAGGCTGACTTCTTGTGTGCTACTTTTCATCCAGGTCTGTAGGTTGCTTATGAAA 1531  
Db 1974 CCATGTCCGATCATCTGTGTGTCACCTTCTCGTCGAGGTGTGCGCGTGGCCTACGAGA 2033  
QY 1532 TCATGCAAACTGCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACT 1591  
Db 2034 TAATGCAGACGATGATCCGGATGACGCGATCGGTTCAAGTCGTGGACGACATATACT 2093  
QY 1592 ATTTTGGAGGCCAAATGCCCAACCAAGATTGCAAGTTTATCTCTCACCACCTCGAACTA 1651  
Db 2094 ACTACGGTGGCCAGAATCGCACATCGCGCGTCTGTTATCGCCCAACAGCGCGCACGC 2153  
QY 1652 AAGAGAAATCCCAATGGAACTCTGAGATATCATTTGGTGTGGCTGGAAACCATTTGGAATG 1711  
Db 2154 ACGAGGATCTACAGCTGCGTGTGTGTGACCTCGTCTCGGTGGCTGGTAAACCATTTGGGATG 2213  
QY 1712 GTTACTCTAAAGGTGTCAACAGAAACTAGGAAACAGGCGCTGTACCCCTTCTCTACAAAG 1771  
Db 2214 GCAATTCNAGGGCAAGATACCCGACCAACCGGGCGGCTATTCCTCTCGTTCAAGG 2273  
QY 1772 TCCGAGAGAAGATAGAAAACAGTCAAAATACCCCTACATATCCTG 1813  
Db 2274 TGGAGGAGAGGTGGACACCGCCAAAGCTGCCTCTCTATGCGG 2315

Search completed: February 2, 2004, 12:33:47  
Job time : 556.385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 12:05:46 ; Search time 120.931 Seconds  
(without alignments)  
7328.922 Million cell updates/sec

Title: US-09-971-773-1

Perfect score: 2008

Sequence: 1 aacagaaactatttctctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1568	78.1	2100	3	US-08-913-805A-9
2	1568	78.1	2100	3	US-09-442-629-9
3	1474.2	73.4	1728	3	US-08-913-805A-1
4	1474.2	73.4	1728	3	US-09-442-629-1
5	79.2	3.9	7218	1	US-08-232-463-14
6	40.8	2.0	1596	4	US-09-134-001C-2083
7	38.4	1.9	1501	4	US-09-173-300-18
8	37.6	1.9	2116	3	US-09-000-041A-1
9	37.6	1.9	5340	4	US-09-627-122-21
10	37.4	1.9	1664976	4	US-08-916-421B-1
11	37.2	1.9	4518	3	US-09-125-287-2
12	36.8	1.8	805	3	US-08-961-083-139
13	36.8	1.8	805	3	US-09-536-784-139
14	36.8	1.8	12127	4	US-08-961-527-148
15	36.2	1.8	12839	3	US-09-125-287-1
16	35.8	1.8	1664976	4	US-08-916-421B-1
17	35.6	1.8	357	4	US-09-134-001C-19
18	35.6	1.8	1368	4	US-09-134-001C-16
19	35.4	1.8	990	4	US-09-134-001C-328
20	35.2	1.8	533	1	US-08-049-283A-1
21	35.2	1.8	858	4	US-09-134-001C-1919
22	35.2	1.8	978	4	US-09-134-001C-110
23	35.2	1.8	1108	4	US-09-702-705-1249
24	35.2	1.8	1108	4	US-09-736-457-1249
25	35.2	1.8	8065	4	US-09-091-952A-6
26	35	1.7	921	4	US-09-328-352-2691
27	34.8	1.7	3487	1	US-08-410-540-4

Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 2, Appli  
Sequence 2383, Ap  
Sequence 12, Appli  
Sequence 3, Appli  
Sequence 1821, Ap  
Sequence 2, Appli  
Sequence 7, Appli  
Sequence 2485, Ap  
Sequence 4, Appli  
Sequence 1471, Ap  
Sequence 105, App  
Sequence 3, Appli  
Sequence 1, Appli

28 34.8 1.7 5361 3 US-08-973-462-2  
29 34.8 1.7 6152 3 US-08-973-462-1  
30 34.2 1.7 1830121 4 US-09-557-884-1  
31 34.2 1.7 1830121 4 US-09-643-990A-1  
32 34 1.7 378 1 US-08-044-547-4  
33 34 1.7 441 1 US-08-044-547-2  
34 34 1.7 699 4 US-09-134-001C-2383  
35 33.8 1.7 999 3 US-09-122-400B-12  
36 33.8 1.7 202001 4 US-09-734-674-3  
37 33.6 1.7 1260 4 US-09-134-001C-1821  
38 33.6 1.7 1835 4 US-09-512-342-2  
39 33.4 1.7 969 4 US-09-080-205-7  
40 33.4 1.7 1383 4 US-09-134-001C-2485  
41 33.4 1.7 2223 1 US-08-257-073-4  
42 33.4 1.7 3255 4 US-09-016-434-1471  
43 33.4 1.7 13158 2 US-08-687-080-105  
44 33.4 1.7 202001 4 US-09-734-674-3  
45 33.4 1.7 1230025 4 US-09-198-452A-1

## ALIGNMENTS

### RESULT 1

US-08-913-805A-9  
; Sequence 9, Application US/03913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCY/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2100 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double

; TOPOLOGY: linear									
; MOLECULE TYPE: cdna									
US-08-913-805A-9									
Query Match									
Best Local Similarity 78.1%; Score 1568; DB 3; Length 2100;									
Matches 1691; Conservative 0; Mismatches 185; Indels 2; Gaps 1;									
Qy	1	AACAGAACTTATTTTCCTGCTGCTACTAGAACACAGAGTACAAATGTTTCCAAATCTT	60						
Db	97	ACAGAGTCTATTCACCTGTCACCTAAGAACAGAGTTACAAATGTTTCAATCTT	156						
Qy	61	TCAGCTCCGAGAGACA--GAAGGAGTTGAAACTCTGAAATCGGGCAGTGAAGTCTG	118						
Db	157	TCAGCTCCAGGACTCCAGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	216						
Qy	119	CTGGGTTGGAATATGCTCATCTTTTCTGCTGGGACCTTATGTTTATATAGGTG	178						
Db	217	CTGGGTTGGAATATGCTCATCTTTTCTGCTGGGACCTTGTCTGTTTATATAGGTG	276						
Qy	179	GTCAATTTGGTTCGAGATAATACCACTGACCAATCTAGCAGAGAACTCTCCAGATTC	238						
Db	277	GTCACTTGTGACGAGTAATGACCATCTGATCACTTAGCCGAGAACTGTCCAGATTC	336						
Qy	239	TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTGAGGAGAATGGCTGAGTCTC	298						
Db	337	TGGCAAGCTTGAACGCTTAAACCAAGCAATGAAGACTTGAGGCGAATGGCCGAATCTC	396						
Qy	299	TCGGAATACCAAGGCCCTATTGATCAGGAGCAGCTACAGGAGAGTCCGTGTTTTAG	358						
Db	397	TCGGATACCAAGGCCCTATTGATCAGGAGCAGCTATAGGAAGAGTACGCGTTTTAG	456						
Qy	359	AAGAACAGCTCTTAAAGGCCAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG	418						
Db	457	AAGAGAGCTTGTAAAGGCCAAGAACAGATTGAAATTTACAGAAACAAAGCTAGGAATG	516						
Qy	419	ATCTGGAAAGGATCATGAAATCTTTAAGGAGGAGATTGAAATCGAGCTAAAGAGTCT	478						
Db	517	GTCTGGGAAGGATCATGAAATCTCTGAGGAGGAGATTGAAATCGAGCTAAAGAGTCT	576						
Qy	479	GGTTTTTCTACAAAGTGAATGAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG	538						
Db	577	GGTTTTTCTACAGAGTGAATGAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG	636						
Qy	539	GACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAAAGTCTATCATGACAGATC	598						
Db	637	GACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAAAGTCTATATGACGATC	696						
Qy	599	TATACCTCAGTCAACAGATGAGCAGAGTGAAGTGGCGGAAAGAAAGCAGAAAGATC	658						
Db	697	TATACCTCAGTCAACAGATGAGCAGAGTGAAGTGGCGGAAAGAAAGCAGAAAGATC	756						
Qy	659	TCACAGAGCTGCTCAGCGGAGAAATACATATCTCAGANTCCCAAGAGCTGCAGCAAG	718						
Db	757	TCACAGAGCTGCTCAGCGGAGAAATACATATCTCAGANTCCCAAGAGCTGCAGCAAG	816						
Qy	719	CCAGAAAGCTGTATGTAATTAACAAGAGCTGTGGCTATGGATGTCAACTCCATCATG	778						
Db	817	CCAGAAAGCTGTGTATTAACAAGAGCTGTGGCTATGGCTGTGCTCCATCATG	876						
Qy	779	TGGTTTACTGCTTCAATGATGCTTATGACCCAGCAGCACTCATCTTGGAAATCTCAGA	838						
Db	877	TGGTTTACTGCTTCAATGATGCTTATGACCCAGCAGCACTCATCTTGGAAATCTCAGA	936						
Qy	839	ATTGGGCTATGCTACTGAGAGCTGGAGACTGTCTTTAGACCTGTAGTGAGCATGCA	898						
Db	937	ATTGGGCTATGCTACTGAGAGCTGGAGACTGTATTTAGGCTGTAAAGTGAAGCATGCA	996						
Qy	899	CAGACAGGCTGGCTCTCCAGTGCACACTGCTGAGGTGAAGTGAAGCAGCAAAATGTTTC	958						
Db	997	CAGACAGATCTGGCATCTCCAGTGCACACTGCTGAGGTGAAGTGAAGCAGCAAAATGTTTC	1056						
Qy	959	AAGTGGTGCAGCTCCCAATTTGAGACAGCCTCCATCTCTGCTCTCTTACCTTTACCTTTGG	1018						

RESULT 2  
US-09-442-629-9  
; Sequence 9, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; UOJIMI, Naofumi  
; SHIBA, Tetsuo  
; YANAGIDANI, Shusaku



1439 TTCCGGGCGGTGATCCTGGATATACACTTTCTCCAGCGTGACTTCCTTGTGTACTT 1498  
1537 TTTCGGAGGTGATCCTGGATATACACTTTCTCTCAGGAGACTTCCTAGTGTACTT 1596  
1499 TTTCATCCAGGTCTGTAGGTTGCTTATGAATCATGCAACACTGATCTGTGCT 1558  
1597 TTTCATCCAGGTCTGTAGGTTGCTTATGAATCATGCAACACTGATCTGTGCT 1656  
1559 CTGCAAACTTTCATCTTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACAACC 1618  
1657 CTGCAAACTTTCATCTTTAGATGACATCTACTATTTTGGAGGCCAAATGCCACAATC 1716  
1619 AGATTGCACTTTATCTCTCAACCACTCGAACTTAAAGAGAAATCCCATGGAACTGGAG 1678  
1717 AAATTGCCATTTATGCTCAACCAACCCGAACCTGCAGATGAAATCCCATGGAACTGGAG 1776  
1679 ATATCATTTGGTGTGGTGGAAACCAATTTGAATGGTTACTCTTAAAGTGTCAACAGAAAC 1738  
1777 ATATCATTTGGTGTGGTGGAAATCATTTGGGATGGCTATTTCTAAAGTGTCAACAGAAAT 1836  
1739 TAGGAAACACGCTGTACCTCTCTCAAAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
1837 TGGGAAGGACGGGCTTATATCTCTCAAAAGTTCGAGAGAGATAGAAACAGTCAAT 1896  
1799 ACCCTACATATCTCTGAAGCTGAAATATAGAGATGGAGTGTAGAGATTTAAACAGAAAT 1858  
1897 ACCCACATATCTCTGAGGCTGAGAAATTAAGCTCAGATGGAAGATTAACGACCAACT 1956  
1859 TAGTTGAGACCATCTCAG 1876  
1957 CAGTTGCAACCAACTCAG 1974

## RESULT 3

US-08-913-805A-1  
; Sequence 1, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:

; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1728 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-913-805A-1

Query Match 73.4%; Score 1474.2; DB 3; Length 1728;  
Best Local Similarity 90.9%; Pred. No. 0;  
Matches 1569; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 100 ATCGGGCATGGACTGTTCTCGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 159  
DB 1 ATCGGGCATGGACTGTTCTCGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60  
QY 160 TTATTGTTTATATAGTGGTCTATTTGGTTCGAGATATGACCACTGACCTTCTAGC 219  
DB 61 TTGCTATTTTACATAGTGGTCTATTTGGTTCGAGATATGACCACTCTGATCATCTAGC 120  
QY 220 AGAGAACTCTCCAAAGTTCCTGCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG 279  
DB 121 CGAGAACTGTCGAAGATTTTGGCAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG 180  
QY 280 AGGAAATGGCTGAGTCTCTCCGAATACCAAGAGCGCTTATGATCAGGGGACAGCTTACA 339  
DB 181 AGGAAATGGCTGAGTCTCTCCGAATACCAAGAGCGCTTATGATCAGGGGACAGCTTACA 240  
QY 340 GGAAGAGTCCGTGTTTATGAGAACACAGCTTGTATGAGGCAAGAACAGATTTGAATTTAC 399  
DB 241 GGAAGAGTCCGTGTTTATGAGAACAGCAATTTATGAGGCAAGAACAGATTTGAATTTAT 300  
QY 400 AAGAAACAAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAA 459  
DB 301 AAGAAACAAAGCTAGGAATGATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTGAA 360  
QY 460 AATGAGCTAAAGAGCTCTGGTCTTCTCAAAAGTGAATTTGAAGAAATTTAAAGAAATTA 519  
DB 361 AATGAGCTAAAGAGCTCTGGTCTTCTCAAAAGTGAATTTGAAGAAATTTAAAGAAATTA 420  
QY 520 GAAGAAACCACTCCAAAGACATGCAAGTGAATTTCTTTTGGATTTAGGACATCATGAA 579  
DB 421 GAAGAAATGAACCTCCAAAGACATGCAAGTGAATTTCTTATCAGATTTGGGACATCATGAA 480  
QY 580 AGGTCTATCATGACAGATCTATCTACTCTCAGTCAAAACAGATGGAGAGGTGAGTGGCG 639  
DB 481 AGGTCTATATGACGGATCTATCTACTCTCAGTCAAAACAGATGGGAGGAGTGGCGT 540  
QY 640 GAAACAAAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAATCTCAGAAAT 699  
DB 541 GAAACGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAATCTCAGAAAT 600  
QY 700 CCCAAGGACTGACAGCAAGCCAGAAAGCTAGTGTATATACAAAGGCTGTGGCTAT 759  
DB 601 CCCAAGGACTGACAGCAAGCCAGAAAGCTAGTGTATATACAAAGGCTGTGGCTAT 660  
QY 760 GGATGTCAACTCCATCATGTGGTCTTACTGCTTATGATTTGCTTATGGCACCACCGCAACA 819  
DB 661 GGCTGTGCTCCATCATGTAGTGTACTGCTTATGATTTGATATGCAATGGCAACCGCAACA 720  
QY 820 CTCACTTTGGAATCTCAGAAATTTGGCGTATGCTACTCGAGATGGGAGAGCTGTGTTTGA 879  
DB 721 CTGCGCTTGGAAATCTCAAAATTTGGCGTACTGCGTACTGGGGAGTGGGAAATCTGTGTTTGA 780  
QY 880 CTTGTAAGTGAGACATGACAGACAGGTCTGGCTCTCCACTGGACACTGGTCAAGGTGAA 939

Db 781 CCTGTAAGTGTGAGCGTGCACAGACAGATCTGGCAGCTCCACTGGACATTTGGTCAAGTGAA 840  
Qy 940 GTGAAGGACAAAATGTTCAAGTGTGTGAGCTGCCATTTGTAGACAGCTCCATCTCGT 999  
Db 841 GTAAAGGACAAAATGTTCAAGTGTGTGAGCTGCCATTTGTAGACAGCTCCATCTCGT 900  
Qy 1000 CCTCTTACTTACCCCTGGCTGTACCAAGAACCTTTGCAGATCGACTCCTGAGAGTCCAT 1059  
Db 901 CCTCATATTTACCCCTGGCTGTCCCAAGAACCTTTGCAGATCGACTTGTACGAGTCCAT 960  
Qy 1060 GGTGATCTGAGTGTGTGGGTATCCAGCTTTGTCAATATCTTATGATCCGTCACACCT 1119  
Db 961 GGTGATCTGAGTGTGTGGGTATCCAGCTTTGTCAATATCTTATGATCCGTCACACCT 1020  
Qy 1120 TGGCTGGAAGGGAATAGAAAGAACCAACCAAGAGCTTTGGCTTCAACATCTCCAGTTAT 1179  
Db 1021 TGGCTGGAAGGGAATAGAAAGAACCAACCAAGAGCTTAGGCTTCAACATCTCCAGTTAT 1080  
Qy 1180 GGAGTCCATGTGACAGCACTGACAAAGTGGAAACAGAGCAGCTTCCATCCCATTTGAG 1239  
Db 1081 GGAGTCCATGTGACAGCACTGACAAAGTGGAAACAGAGCAGCTTCCATCCCATTTGAG 1140  
Qy 1240 GAATACATGTTACAGTTGAAGACATTTTTCAGCTTCTCGAACGAGCAATGAAGTGGAT 1299  
Db 1141 GAATACATGTTACAGTTGAAGACATTTTTCAGCTTCTCGAACGAGCAATGAAGTGGAT 1200  
Qy 1300 AAAAAAGAGTGTATCTGCCCACTGATGACCTCTTTTAAAGAGGCAAGACAAAG 1359  
Db 1201 AAAAAAGAGTGTATTTGGCCACAGATGACCTCTTTTAAAGAGGCAAGACAAAG 1260  
Qy 1360 TACTCCATATGATTTTATGATGAATCTATTTCTTGGTCACTGGACTACACAA 1419  
Db 1261 TACCCAGTTATGATTTTATGATGAATCTATTTCTTGGTCACTGGACTACATAAT 1320  
Qy 1420 CGATACAGAAAATCTCTCGGGCGTGTATCTGGATATACACTTTCTCTCCAGGCT 1479  
Db 1321 CGATATACAGAAAATCTCTCGGGCGTGTATCTGGATATACACTTTCTCTCCAGGCA 1380  
Qy 1480 GACTTCTTGTGTACTTTTTCATCCAGCTCTGTAGGTTGCTTATGAATCATGCA 1539  
Db 1381 GACTTCTTGTGTACTTTTTCATCCAGCTCTGTAGGTTGCTTATGAATCATGCA 1440  
Qy 1540 ACATGCACTCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1599  
Db 1441 GCGTGCATCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1500  
Qy 1600 GGCCAAAATGCCCAACACAGATTCAGTTTATCTCAACCACTTCGAATTAAGAGAGAA 1659  
Db 1501 GGCCAAAATGCCCAACACAAATTTGCTTATCTCAACCACTTCGAATTAAGAGAGAA 1560  
Qy 1660 ATCCCATGGAACTGGAGATATCTTGGTGGCTGGAAACCATTTGAATGGTTACTCT 1719  
Db 1561 ATCCCATGGAACTGGAGATATATTTGGTGGCTGGAAATCACTGGGATGGCTATCTCT 1620  
Qy 1720 AAAGTGTCAACAGAAAACTAGGAAAAAGAGGCTGTACCTTCTCAAAAGTCCGAGAG 1779  
Db 1621 AAAGTGTCAACAGAAAACTGGAGAGGAGGCGCTATATCTCTCAAAAGTTCGAGAG 1680  
Qy 1780 AAGATAGAAAAGTCAAAATACCTTACATATCTTGAAGCTGAAAAATA 1826  
Db 1681 AAGATAGAAAAGTCAAAATACCTTACATATCTTGAAGCTGAAAAATA 1727

RESULT 4

US-09-442-629-1

Sequence 1, Application US/09442629

Patent No. 6291219

GENERAL INFORMATION:

APPLICANT: TANIGUCHI, Naoyuki

UOZUMI, Naofumi

SHIBA, Tetsuo

YANAGIDANI, Shusaku

TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

CITY: Washington

STATE: DC

COUNTRY: US

ZIP: 20036

MEDIUM TYPE: 3+ Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2

SOFTWARE: WordPerfect 6.1 Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/442.629

FILING DATE: 18-Nov. 6291219-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/913.805A

FILING DATE: 7 JAN 1998

APPLICATION NUMBER: PCT/JP97/00171

FILING DATE: 23 JAN 1997

APPLICATION NUMBER: JP 192260

FILING DATE: 22 JUL 1996

APPLICATION NUMBER: JP 162813

FILING DATE: 24 JUN 1996

APPLICATION NUMBER: JP 161648

FILING DATE: 21 JUN 1996

APPLICATION NUMBER: JP 10365

FILING DATE: 24 JAN 1996

ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.

REGISTRATION NUMBER: 39,048

REFERENCE/DOCKET NUMBER: 2356/3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-429-1776

TELEFAX: 202-425-0796

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1728 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-442-629-1

Query Match 73.4%; Score 1474.2; DB 3; Length 1728;  
Best Local Similarity 90.9%; Pred. No. 0;  
Matches 1569; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 100 ATGCGGCGCATGGAGCTGTTCTCGGCGTTGGATTATGCTCATTTTTCCTGGGGGACC 159  
Db 1 ATGCGGCGCATGGAGCTGTTCTCGGCGTTGGATTATGCTCATTTTTCCTGGGGGACC 60  
Qy 160 TTATTTGTTTATATAGTGTGTTGTTGTTTTCGAGATATGACCCCTGACCATTTCTAGC 219  
Db 61 TTGCTATTTTACATAGTGTGTTGTTGTTTTCGAGATATGACCCCTGACCATTTCTAGC 120  
Qy 220 AGAGAACTCTCCAAAGATCTTTCGAAAGCTGGAGCGCTTAAACAAACAAATGAAGACTTG 279  
Db 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAAGCGCTTAAACAAACAAATGAAGACTTG 180  
Qy 280 AGGAGATGGCTGAGTCTCTCCGAATACCGAAGCGCTTATTTGATCAGGGGACAGCTACA 339  
Db 181 AGGAGATGGCTGAGTCTCTCCGAATACCGAAGCGCTTATTTGATCAGGGGACAGCTACA 240  
Qy 340 GGAAGAGTCCGTTGTTTAGAAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAATATAC 399  
Db 241 GGAAGAGTCCGTTGTTTAGAAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAATAT 300  
Qy 400 AAGAAACAAAGCTAGGAATGATCTCGGAAAGGATCATGAAATCTTTAAGAGAGGAGATTGAA 459  
Db 301 AAGAAACAAACTAAAAATGGTCCAGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360

QY 460 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTA 519  
Db 361 AATGAGCTAAGAGCTCTGTTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTA 420  
QY 520 GAAGAAAGCACTCAAGACATCGAGTGAATTTCTTTGGATTTAGGACATCATGAA 579  
Db 421 GAAGAAATGAATCTCAAGACATCGAGTGAATTTCTTTGGATTTAGGACATCATGAA 480  
QY 580 AGGTCTATCATGACAGATCTTACTACTCAAGACATCGAGTGAATTTAGGACATCATGAA 639  
Db 481 AGGTCTATTAAGACAGATCTTACTACTCAAGACATCGAGTGAATTTAGGACATCATGAA 540  
QY 640 GAAAGAGAGCCAAAGATCTGACAGAGCTGCTGACAGAGTGAATTTAGGACATCATGAA 699  
Db 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGCTGACAGAGTGAATTTAGGACATCATGAA 600  
QY 700 CCAAGAGAGTGAAGAAAGCCAAAGATCTGATGATGATGATGATGATGATGATGATGAT 759  
Db 601 CCAAGAGAGTGAAGAAAGCCAAAGATCTGATGATGATGATGATGATGATGATGATGAT 660  
QY 760 GGATGTCATCACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 819  
Db 661 GGCTGTCATCACTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
QY 820 CTGATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGAGAGTGGAGACTGTGTTTGA 879  
Db 721 CTGCGCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGAGAGTGGAGACTGTGTTTGA 780  
QY 880 CTTGTAAGTGAATGATGACAGAGTCTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 939  
Db 781 CTTGTAAGTGAATGATGACAGAGTCTGGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
QY 940 GTGAGGACAAAGATTTCAAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999  
Db 841 GTGAGGACAAAGATTTCAAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
QY 1000 CTTCTTACTTACCTTGGCTGTACAGAAAGCTTGGCTTCAAGATCTGCTGCTGCTGCTGCT 1059  
Db 901 CTTCTTACTTACCTTGGCTGTACAGAAAGCTTGGCTTCAAGATCTGCTGCTGCTGCTGCT 960  
QY 1060 GGTGATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119  
Db 961 GGTGATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
QY 1120 TGGCTGGAAGGGAATAGAGAAACACCAAGAGCTTGGCTTCAAGATCTGCTGCTGCTGCT 1179  
Db 1021 TGGCTGGAAGGGAATAGAGAAACACCAAGAGCTTGGCTTCAAGATCTGCTGCTGCTGCT 1080  
QY 1180 GGAGTCCATGTACAGCGACCTGACAAAGTGGGAAACAGAGAGCTTCCATCCCATTTGAG 1239  
Db 1081 GGAGTCCATGTACAGCGACCTGACAAAGTGGGAAACAGAGAGCTTCCATCCCATTTGAG 1140  
QY 1240 GAATACATGTTACGTTTGAAGAAATTTTTCAGCTTCTGAAACGAGAGTGAATGAGT 1299  
Db 1141 GAATACATGTTACGTTTGAAGAAATTTTTCAGCTTCTGAAACGAGAGTGAATGAGT 1200  
QY 1300 AAAAAAGAGTGTATCTGCGCACTGATGACCTTCTTTGTTAAAGAGGCAAGAAAGCAAG 1359  
Db 1201 AAAAAAGAGTGTATCTGCGCACTGATGACCTTCTTTGTTAAAGAGGCAAGAAAGCAAG 1260  
QY 1360 TACTCCAAATTTAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1419  
Db 1261 TACTCCAAATTTAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
QY 1420 CGATACAGAAATTTCACTTGGGGGCTGATCTGATGATGATGATGATGATGATGATGATGAT 1479  
Db 1321 CGATACAGAAATTTCACTTGGGGGCTGATCTGATGATGATGATGATGATGATGATGATGAT 1380  
QY 1480 GACTTCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1539  
Db 1381 GACTTCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440

QY 1540 AACTGTCATCTGATGCTCTGCAAACTTCATCTTTTAGATGATGATGATGATGATGATGAT 1599  
Db 1441 GCGTGTGATCTGATGCTCTGCAAACTTCGATCTTTTGGATGATGATGATGATGATGATGAT 1500  
QY 1600 GGCCAAATGTCACCAACAGATTTGATCTTCTCACCAACCTCGAACTTAAAGAGGAA 1659  
Db 1501 GGCCAAATGTCACCAACAGATTTGATCTTCTCACCAACCTCGAACTTAAAGAGGAA 1560  
QY 1660 ATCCCCATGGAACCTGGAGATATCATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1719  
Db 1561 ATCCCCATGGAACCTGGAGATATTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620  
QY 1720 AAAGTGTCAACAGAAACTAGGAAACAGGCTGTACCTTCTTCAAAAGTCCGAGAG 1779  
Db 1621 AAAGTGTCAACAGAAACTAGGAAACAGGCTGTATATCTTCTTCAAAAGTCCGAGAG 1680  
QY 1780 AAGATAGAAACAGTCAAAATACCTTACATATCTTGAAGCTGAAATAA 1826  
Db 1681 AAGATAGAAACAGTCAAAATACCTTACATATCTTGAAGCTGAAATAA 1727

## RESULT 5

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F18  
; US-08-232-463-14

Query Match 3.9%; Score 79.2; DB 1; Length 7218;  
Best Local Similarity 6.6%; Pred. No. 1e-12;  
Matches 27; Conservative 236; Mismatches 149; Indels 0; Gaps 0;





## US-09-000-041A-1

Query Match 1.9%; Score 37.6; DB 3; Length 2116;  
Best Local Similarity 49.6%; Pred. No. 0.75;  
Matches 126; Conservative 0; Mismatches 124; Indels 4; Gaps 1;  
  
QY 379 AAAGAACAGATTGAAATTTACAGAAACAAAGCTAGCAATGATCTCGGAAAGATCATGAA 438  
DB 1507 AATGAAACAAATGGACCTGGCTATCCCAAGCCAATCAAGTCTCGAAAGGATCCCTGA 1566  
  
QY 439 ATCTTAAGGAGGAGATTGAAATCGAGCTAAAGAGCTCTGTTTTTTCTACAAG---- 494  
DB 1567 ATCTCCTCAGGAGGATTGTTACACAAAGAAATGCTTTTACGTTATTTCTCAAGAGAA 1626  
  
QY 495 TGAATTGAAGAAATTAAGAAATTAAGAGGAAACCAACTCCAAAGACATGAGATGAAAT 554  
DB 1627 GGAGTATTGGAATTTCAACACCAATCACTCAAGTAGAATCTGACATCCAAGATCAT 1686  
  
QY 555 TCTTTTGGATTAGACATCATGAAAGGTCTATCATGACAGATCTATCTACCTCAGTCA 614  
DB 1687 CCTCAAGATTTTATGGGCTGTGATGGACCAACAGACAGAGTTAAAGAGGACACAGCCC 1746  
  
QY 615 AACAGATGGAGCAG 628  
DB 1747 ACCAGATGATGATG 1760

## RESULT 9

US-09-627-122-21/c  
; Sequence 21, Application US/09627122  
; Patent No. 6472521  
; GENERAL INFORMATION:  
; APPLICANT: Uhlmann, Eugen  
; APPLICANT: Greiner, Beate  
; APPLICANT: Unger, Eberhard  
; APPLICANT: Gothe, Gislinde  
; APPLICANT: Schwerdel, Marc  
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN EGS  
; FILE REFERENCE: 02481.1678  
; CURRENT APPLICATION NUMBER: US/09/627,122  
; CURRENT FILING DATE: 2000-07-27  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 5340  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-09-627-122-21

Query Match 1.9%; Score 37.6; DB 4; Length 5340;  
Best Local Similarity 54.3%; Pred. No. 1.2;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
  
QY 426 AAAGATCATGAAATCTTAAGGAGGAGGATTGAAATGGAGCTAAAGAGCTCTGTTTTT 485  
DB 3851 AAAAGAGGACAGCTGAAACGCTTATATAAAATCGAGTAGAAGTCAATTCGATTTTT 3792  
  
QY 486 TCTCAAGTGAATTTGAAGAAATTAAGAAATTAAGAAACGAACTCCAAAGCATGC 545  
DB 3791 ACTATAACATTAATTTATAAAGATATAAATATGTAGGAGAAATTAACCAAAATAGGA 3732  
  
QY 546 AGATCAAAATCTTTTGGATT 565  
DB 3731 AAATTAATTTAGTTGATT 3712

## RESULT 10

US-08-916-421B-1/c  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.

; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
; Patent No. 6503729  
; TITLE OF INVENTION: jannaschii  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B  
; CURRENT FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/024,428  
; PRIOR FILING DATE: 1996-08-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1664976  
; TYPE: DNA  
; ORGANISM: Methanococcus jannaschii  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (28222)..(28222)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (28257)..(28258)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84773)..(84773)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84808)..(84808)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84812)..(84812)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98120)..(98120)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98159)..(98159)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98239)..(98239)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98266)..(98266)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98343)..(98343)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (103998)..(103998)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (148948)..(148948)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (163385)..(163385)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191989)..(191989)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191995)..(191995)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (231980)..(231980)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (234187)..(234187)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (234220)..(234220)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (234814)..(234814)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature

LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559167)..(559167)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (600992)..(600992)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (622708)..(622708)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657081)..(657081)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657203)..(657203)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (713652)..(713652)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779576)..(779576)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)

OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491)..(1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091)..(1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020)..(1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912)..(1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734)..(1603734)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998)..(1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664854)..(1664854)  
OTHER INFORMATION: n equals a, t, c, or g  
US-08-916-421B-1

Query Match 1.0%; Score 37.4; DB 4; Length 1664976;  
Best Local Similarity 47.3%; Pred. No. 22;  
Matches 113; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
Qy 351 TGTTTTGAAGAACAGCTTGTAAAGGCCAACAGAGATTCAGAAATTAACAAGAACAAAGC 410  
Db 331793 TGTCTTAAGACACCATTTGGATGTATAGGCATTTAAATCAGCTGAAGGACATCC 331734  
Qy 411 TAGGAATGATCTGGGAAGGATCATGAATCTTAAGGAGGAGGATTTGAAATGGAGCTAA 470  
Db 331733 TTTGAGTGAGTTGTCAAGAAAGATTGTAGCTAAGGAGAAAGAAAGAGAGAGA 331674  
Qy 471 AGAGCTCTGGTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTAAGAAAGAGA 530  
Db 331673 ATCAAAATTCCTACTTATTAAACCCAGAACACAGAGGAAATTTGAATTTGAATGA 331614  
Qy 531 ACTCCAAAGACATGCAGATGAATTTCTTTTGGATTTAGGACATCATGAAAGCTCTATCA 589  
Db 331613 CATAAATATAATTAAGAGATTTAGCTTTAGCTTAACATGAATTTGGGATTA 331555

## RESULT 11

US-09-125-287-2/c  
Sequence 2, Application US/09125287B  
Patent No. 6114602  
GENERAL INFORMATION:  
APPLICANT: BARG, Rivka  
APPLICANT: SALTS, Tehlan  
TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAPI IN PLANTS  
FILE REFERENCE: INTRON GENETIC PARTHENOCAPI IN PLANTS  
CURRENT APPLICATION NUMBER: US/09/125.287B  
CURRENT FILING DATE: 1998-11-09  
EARLIER APPLICATION NUMBER: PCT/IL97/00051  
EARLIER FILING DATE: 1997-02-13  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 2  
LENGTH: 4518  
TYPE: DNA  
ORGANISM: TPRP-F1 PROMOTOR  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (6)..(191)  
OTHER INFORMATION: "n"s are any nucleic residue  
US-09-125-287-2



APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340F1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 148:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12127 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-148

Query Match 1.8%; Score 36.8; DB 4; Length 12127;  
Best Local Similarity 50.6%; Pred No. 3;  
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
Qy 272 AAGACTTCAGGAGGATGGCTGAGTCTCCGAATACCAAGAGGCCCTTATCATCAGGGGA 331  
Db 8181 AAGAAATTACAAACAAATTCATTTATCAAAAAGTTGATGAAACTTTTAAACAAGAGA 8240  
Qy 332 CAGCTACGAGGAGAGTCGCTGTTTAGAAGAACAGCTTGTAAAGCCCAAGAACAGATTG 391  
Db 8241 ATTTGGAAGAACTCTTAAATCTCTAAATGATCTTTGTTGATAAATATCAAAAACAAATCG 8300  
Qy 392 AAAATTACAGAAACAAAGCTAGGATGATCTGGGAAAGGATCATGAAATCTTAAGG 447  
Db 8301 AACTTTGAAGAAAGAGAGAAAGAAAGCTGCTGAAAAAGCTGCTGAAAAAGCAAG 8356

## RESULT 15

US-09-125-287-1/c  
Sequence 1, Application US/09125287B  
Patent No. 6114602  
GENERAL INFORMATION:  
APPLICANT: BARG, Rivka  
APPLICANT: SALT, Tehiam  
TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAPI IN PLANTS  
FILE REFERENCE: INTRON GENETIC PARTHENOCAPI IN PLANTS  
CURRENT APPLICATION NUMBER: US/09/125,287B  
CURRENT FILING DATE: 1998-11-09  
EARLIER APPLICATION NUMBER: PCT/IL97/00051  
EARLIER FILING DATE: 1997-02-13  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 12839  
TYPE: DNA

ORGANISM: TPRP-F1 GENOMIC CLONE  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (5) (11322)  
OTHER INFORMATION: "n"s are any nucleic residue  
US-09-125-287-1  
Query Match 1.8%; Score 36.2; DB 3; Length 12839;  
Best Local Similarity 22.5%; Pred. No. 4.6;  
Matches 49; Conservative 67; Mismatches 102; Indels 0; Gaps 0;  
Qy 1586 TCTACTATTTTGGAGGCCAAAATCCCAACACAGATTGCGATTATCTCCTCACCACCTC 1645  
Db 218 TCGCTATTTGGGATGCATTATTHDSVANAASNAATANNAYGNTYADASNASHVHN 159  
Qy 1646 GAACATAAGAGGAATCCCATGGAACCTGGAGATATCATTTGGTGGCTGGAACCAATT 1705  
Db 158 ADASRTNTNSYYAKYSACDAYHSNHNATANNATNTYADASNASSYAAHVNHN 99  
Qy 1706 GGAATGGTTACTCTAAAGGTGTCAACAGAAAACCTAGGAAAACAGGGCTGTACCTTCT 1765  
Db 98 HGCATDAYYTHDSVASTWKYCNHGNAGNASDAYYTHDSVAANSTWKYCNHNNANCYAKYDA 39  
Qy 1766 ACAAGTCCGAGAGAGATAGAAAACAGTCAATACCTT 1803  
Db 38 GNASDANCYAKYDAVNCNHAGNNGKNCYADAGNASVT 1  
Search completed: February 2, 2004, 18:41:04  
Job time : 152.431 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 12:16:51 ; Search time 702.477 Seconds  
(without alignments)  
10417.799 Million cell updates/sec

Title: US-09-971-773-1  
Perfect score: 2008  
Sequence: 1 aacagaaactatttctg.....gctgcccctcaagcccatg 2008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	100.0	2008	11 US-09-971-773-1	Sequence 1, Appli
2	1584	78.9	1728	11 US-09-971-773-2	Sequence 2, Appli
3	1574.2	78.4	3007	15 US-10-106-698-381	Sequence 381, App
4	1569.6	78.2	2100	9 US-09-839-136-9	Sequence 9, Appli
5	1472.6	73.3	1728	9 US-09-839-136-1	Sequence 1, Appli
6	976.4	48.6	979	11 US-09-971-773-6	Sequence 6, Appli
7	893.2	44.5	922	11 US-09-971-773-7	Sequence 7, Appli
8	368.2	18.3	422	11 US-09-918-995-33468	Sequence 33468, A
9	302	15.0	9196	11 US-09-971-773-3	Sequence 3, Appli
10	285.6	14.2	503	9 US-09-864-761-13359	Sequence 13359, A
11	218.8	10.9	551	9 US-09-864-761-13292	Sequence 13292, A
12	213.4	10.6	248	9 US-09-864-761-18181	Sequence 18181, A
13	213.4	10.6	248	9 US-09-864-761-30284	Sequence 30284, A
14	149	7.4	481	13 US-10-125-968-274	Sequence 274, App
15	136.4	6.8	384	9 US-09-864-761-1421	Sequence 1421, Ap

16	117.4	5.8	151	12	US-10-242-535A-52289	Sequence 52289, A
17	112.8	5.6	419	11	US-09-918-995-33106	Sequence 33106, A
18	103.4	5.1	129	9	US-09-864-761-29920	Sequence 29920, A
19	53.8	2.7	65	13	US-09-908-975-23971	Sequence 23971, A
20	46.2	2.3	60	13	US-09-908-975-4992	Sequence 4992, Ap
21	42.8	2.1	1137	15	US-10-184-644-518	Sequence 518, App
22	42.8	2.1	1137	15	US-10-184-634-518	Sequence 518, Appli
23	40.8	2.0	3673778	13	US-10-312-841-2	Sequence 2, Appli
24	39	1.9	581	10	US-09-789-054A-11	Sequence 11, Appli
c 25	38.8	1.9	1485	10	US-09-764-877-3640	Sequence 3640, Ap
c 26	38.8	1.9	1485	11	US-09-764-891-5860	Sequence 5860, Ap
c 27	38.8	1.9	1485	12	US-10-242-515-3640	Sequence 3640, Ap
28	38.4	1.9	1501	14	US-10-027-450-18	Sequence 18, Appli
29	38.2	1.9	11872	10	US-09-764-847-1390	Sequence 1390, Ap
30	38.2	1.9	11872	15	US-10-092-154-1390	Sequence 1390, Ap
c 31	38.2	1.9	3673778	13	US-10-312-841-1	Sequence 1, Appli
32	38	1.9	716	13	US-10-140-472-96	Sequence 96, Appli
33	38	1.9	716	13	US-10-141-761-96	Sequence 96, Appli
34	38	1.9	716	13	US-10-142-885-96	Sequence 96, Appli
35	38	1.9	716	13	US-10-158-790-96	Sequence 96, Appli
36	38	1.9	716	13	US-10-137-871-96	Sequence 96, Appli
37	38	1.9	716	13	US-10-140-805-96	Sequence 96, Appli
38	38	1.9	716	13	US-10-140-864-96	Sequence 96, Appli
39	38	1.9	716	13	US-10-140-923-96	Sequence 96, Appli
40	38	1.9	716	13	US-10-141-756-96	Sequence 96, Appli
41	38	1.9	716	13	US-10-141-759-96	Sequence 96, Appli
42	38	1.9	716	15	US-10-123-155-96	Sequence 96, Appli
43	38	1.9	716	16	US-10-146-731-96	Sequence 96, Appli
c 44	38	1.9	941	13	US-10-140-472-464	Sequence 464, App
c 45	38	1.9	941	13	US-10-141-761-464	Sequence 464, App

## ALIGNMENTS

RESULT 1  
US-09-971-773-1  
; Sequence 1, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: NO. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patent in ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2008  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-1

Query Match 100.0%; Score 2008; DB 11; Length 2008;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AACGAACTATTATTTCTGTGTGCTAAGTACAGTACAAATGTTTCCAAATCTT 60  
Db 1 AACGAACTATTATTTCTGTGTGCTAAGTACAGTACAAATGTTTCCAAATCTT 60  
Qy 61 TGAGCTCCGAGACAGACAGAGGAGTTTCAAACTCTGAAATCGGGGATCGACTGGTTC 120





```

; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971,773
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: JP 2000-308526
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/268,926
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-971-773-2

```

Query Match	78.9%	Score 1584	DB 11	Length 1728
Best Local Similarity	94.8%	Prod. No. 0		
Matches 1638	Conservative 0	Mismatches 90	Indels 0	Gaps 0
Qy	100	ATCGCGGCATGCATCGTGTCTCGCGTTCGGATTAATGCTCATTTCTTTTTCCTCGCGGGAC	159	
Db	1	ATCGCGGCATGCATCGTGTCTCGCGTTCGGATTAATGCTCATTTCTTTTTCCTCGCGGGAC	60	
Qy	160	TTATTGTTTATATAGTGGTCAATTTGGTTCGAGATAATGACCACTTACCATTTCTAGC	219	
Db	61	TTGTTATTTATATAGTGGTCAATTTGGTTCGAGATAATGACCACTTACCATTTCTAGC	120	
Qy	220	AGAACTCTCCAAGATCTTTGCAAGCTGGAGCGCTTAAACAAACAAATTAAGACATTG	279	
Db	121	AGAACTCTCCAAGATCTTTGCAAGCTGGAGCGCTTAAACAAACAAATTAAGACATTG	180	
Qy	280	AGGAGATGGCTGAGTCTCCGGAATACAGAGCCCTATTGATCAGGGGACACTACA	339	
Db	181	AGGCGAATGGCTGAGTCTCTCCGAATACAGAGGCCCTATTGATCAGGGGACACTACA	240	
Qy	340	GGAGAGTCCGTGTTTATAGAGAACAGCTGTGTTAAGGCCAAGAACAGATTGAAATTTAC	399	
Db	241	GGAGAGTCCGTGTTTATAGAGAACAGCTGTGTTAAGGCCAAGAACAGATTGAAATTTAC	300	
Qy	400	AAGAAACAGCTAGGAATGATCTGGGAAAGGATCATGAATCTTAAAGAGGAGGATTGAA	459	
Db	301	AAGAAACAGCTAGGAATGATCTGGGAAAGGATCATGAATCTTAAAGAGGAGGATTGAA	360	
Qy	460	AATCGAGCTAAGAGAGCTCTGGTTTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTA	519	
Db	361	AATCGAGCTAAGAGAGCTCTGGTTTTTCTCAAAAGTGAATTAAGAAATTAAGAAATTA	420	
Qy	520	GAAGGAAACGAACCTCCAAGACATGCAGATGAAATTTCTTTGGATTAGGACATCATGAA	579	
Db	421	GAAGGAAATGAACCTCCAAGACATGCAGATGAAATTTCTTTGGATTAGGACATCATGAA	480	
Qy	580	AGGTCTATCATGACAGATCTATACCTCAGTCAACAGATGGAGCAGGTGAGTGGCGG	639	
Db	481	AGGTCTATCATGACAGATCTATACCTCAGTCAACAGATGGAGCAGGTGAGTGGCGG	540	
Qy	640	GAAAAGAAAGCCAAAGATCTCAGACAGCTGTGTCAGCGGAGAAATAACATATCTCGAGAT	699	
Db	541	GAAAAGAAAGCCAAAGATCTCAGACAGCTGTGTCAGCGGAGAAATAACATATCTCGAGAT	600	
Qy	700	CCCAGGACCTCAGCAAGCCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTAT	759	
Db	601	CCTAAGGACCTCAGCAAGCCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTAT	660	
Qy	760	GGATGTCACCTCATATGTTGGTTTATGCTTTCATGATTGCTTATGGCACCCAGCGGAACA	819	
Db	661	GGTTGTCAACTCCATCAGCTGGTCTACTGTTTTCATGATTGCTTATGGCACCCAGCGGAACA	720	
Qy	820	CTCATCTTGGAACTCTCAAAATTTGGCGCTATGCTACTGGAGGATGGAGACTGTGTTTAGA	879	
Db	721	CTCATCTTGGAACTCTCAAAATTTGGCGCTATGCTACTGGAGGATGGAGACTGTGTTTAGA	780	
Qy	880	CCTGTAAGTGAGACATGCACAGACAGGCTGTGGCTCTCCACTGGACACTGGTTCAGGTGAA	939	

Qy	1	AACAGAAAC	TTATTTT	CCTGTGTGGCT	AAC	CTAGAAC	CCAGAGT	ACAAT	GTGTTT	CCAATTTCTT	60
Db	319	AACAGAAAGT	CTTATTT	CACCTGTGCA	CTAACT	AGAAAC	CAGAGT	TAACA	ATGTTT	TTTCAATTTCTT	378
Qy	61	TGAGCTCCG	AGAGACA	- -	GAAGGGAG	TGAAA	CTCTG	AAAAT	CGGGCA	TGAGCTGGTT	118
Db	379	TGAGCTCCG	AGACTY	CAGGNA	GTGAGT	TGAAA	ATCTG	AAAAT	CGGGCC	TGAGCTGGTT	438
Qy	119	CTTGGCGTT	TGGAAT	TATGCT	CATCTCTTTT	TGCT	GGGGA	CTTAT	TGTTTT	TATATAGGTG	178
Db	439	CTTGGCGTT	TGGAAT	TATGCT	CATCTCTTTT	TGCT	GGGGA	CTTAT	TGTTTT	TATATAGGTG	498
Qy	179	GTCAATTT	TGGTT	TCGAGAT	TAATG	ACCAC	CTGAC	CAATCT	TAGCAG	AGAACTCT	238
Db	499	GTCAATTT	TGGTAC	GAGAT	TAATG	ACCAT	CTGTAT	CACTCT	TAGCCG	AGAACTGT	558
Qy	239	TTGCAAA	AGCTGG	AGCGCT	TAAAA	CAACA	AAATGA	AGACTT	TGAGG	AGAAATG	297
Db	559	TGGCAAA	AGCTTGA	ACGCTT	AAAA	CAGC	AGATGA	AGACTT	TGAGG	CGGAATG	618
Qy	298	CTCCGA	ATAC	CAGAA	AGGCCCT	TATTTG	ATCAG	GGGA	CAGCT	TACAGAA	357
Db	619	CTCCGA	ATAC	CAGAA	AGGCCCT	TATTTG	ATCAG	GGGCC	CAGCT	TAGAA	678
Qy	358	GAAGAA	CAGCTTG	TTAAG	GGCCAA	AGAA	CAGAT	TGAAA	ATTA	CAAGAA	417
Db	679	GAAGAG	CAGCTTG	TTAAG	GGCCAA	AGAA	CAGAT	TGAAA	ATTA	CAAGAA	738
Qy	418	GATCTGG	GAAGG	ATCAT	GAAATCTT	TAA	GAGG	AGGAT	TGAAA	ATGAGCT	477
Db	739	GGTCTGG	GAAGG	ATCAT	GAAATCTT	TAA	GAGG	AGGAT	TGAAA	ATGAGCT	798
Qy	478	TGGTTTTT	TTCTA	CAAGTGA	TTTGA	AGAAAT	TAA	AGAAAT	TAG	AGAGAA	537
Db	799	TGGTTTTT	TTCTA	CAGATGA	TTTGA	AGAAAT	TAA	AGAAAT	TAG	AGAGAA	858
Qy	538	AGACAT	GCAGAT	GAAATTTCTTTT	TGGATTT	TAGG	ACAT	CAT	GAAAG	GTCTAT	597
Db	859	AGACAT	GCAGAT	GAAATTTCTTTT	TGGATTT	TAGG	ACAT	CAT	GAAAG	GTCTAT	918
Qy	598	CTATAC	TACCT	CAGTCA	AA	CHAG	TGG	AGC	AGGTG	GGGGAAAA	657
Db	919	CTATAC	TACCT	CAGTCA	GA	CA	GAT	TGG	AGC	AGGTG	978
Qy	658	CTGAC	CAGAGCT	TGGTCC	AGCG	AGAAAT	TA	CA	TATCT	GCA	717

Db 2059 TTGGGAAGCGGCGCTATATCCCTCTCAAAAGTTTCGAGAGAAATAGAAACGGTCAAG 2118  
Qy 1798 TACCTCATATCTGAAGCTGAAATAAGATGAGTGAAGATTA- - - - - 1848  
Db 2119 TACCCACATATCTGAGGCTGAGAAATAAAGCTCAGATGAAGAGATAAACGACCAAC 2178  
Qy 1849 - - - - - CAACAGAAATTTAGTTTCAGACCATCTCAGCCAGCAGAGAACCCAG- - - - - 1893  
Db 2179 TCAGTTCGACCAACTCAGTTTCAAAACCATTTCCAGCCAACTGTAGATGAAGAGGCTCTG 2238  
Qy 1894 - - - - - ACTAACATATGTTTCATTTGACAGACATGCTCCGACCAAGAGCAAGTGGGAACCC 1948  
Db 2239 ATCTAACAAAATAAGGTTATATGATAGATATCTCTCAGCACCAAGAGCAGCTGGGAACTG 2298  
Qy 1949 TCAGATGCTGC- ACTGGTGGACGCTCTTTGTGAAGGCTGCTGCTCCCTCAAGCCCAT 2007  
Db 2299 ACATAGGCTTCAATTTGGTGAATTCCTCTTTAAACAAGGGCTGCAATGCCCTCATACCCAT 2358  
Qy 2008 G 2008  
Db 2359 G 2359

## RESULT 4

US-09-839-136-9  
; Sequence 9, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 2100  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (198)...(1925)  
US-09-839-136-9

Query Match 78.2%; Score 1569.6; DB 9; Length 2100;  
Best Local Similarity 90.1%; Pred. No. 0;  
Matches 1692; Conservative 0; Mismatches 184; Indels 2; Gaps 1;  
Qy 1 AACGAAATCTATTTCTGTGTGCTTAACCTAGAACACAGAGTACAAATGTTTCCAAATCTT 60  
Db 97 AACGAAAGTCTATTCACCTGTGCACTAATAAGAAACAGAGTTACAAATGTTTCAATCTT 156  
Qy 61 TGAGCTCCGAGAGACA- - GAAGGGAGTTGAAACTCTGAAATCGGGCATGGACTGGTT 118  
Db 157 TGAGCTCCAGGACTCCAGGGAAGTGAGTTGAAATCTGAAATCGGGCATGGACTGGTT 216  
Qy 119 CTGGCGTTGATATGCTCATTCTTTTGGCTGGGGACCTTATTTTATATAGTG 178  
Db 217 CTGGCGTTGATATGCTCATTCTTTTGGCTGGGGACCTTGTCTGTTTATATAGTG 276  
Qy 179 GTCAATTTGTTTCAGATATACCACTGACCATTTCTAGCAGAGAACTCTCCAGATTC 238  
Db 277 GTCACTTGTGACGATATATGACCTCTGATCATCTAGCCGAGACTGTCAGATTC 336  
Qy 239 TTGCAAGCTGAGCGCTTAAACCAACAAATGAAGACTTGAAGGAGATGGCTGAGTCTC 298  
Db 337 TGGCAAGCTTGAACGCTTAAACACAGCAGATGAAGACTTGAAGGAGATGGCGAATCTC 396

Qy 299 TCCGAATACCAAGAGCCCTATTGATCAGGGGACAGCTACAGGAAGAGTCCGTTTGTAG 358  
Db 397 TCCGATACCAAGAGCCCTATTGATCAGGGCCAGCTATAGGAAGAGTAGCGCTTTAG 456  
Qy 359 AAGAACAGCTTGTAAAGGCCAAAGAACAGATTGAATAATTACAAGAAACAAGCTAGGATG 418  
Db 457 AAGAGCAGCTTGTAAAGGCCAAAGAACAGATTGAATAATTACAAGAAACAAGCTAGGATG 516  
Qy 419 ATCTGGGAAGGATCATGAATCTTAAAGGAGGAGATTGAAATGGAGCTTAAAGAGCTCT 478  
Db 517 GTCTGGGAAGGATCATGAATCTTGAAGGAGGAGATTGAAATGGAGCTTAAAGAGCTCT 576  
Qy 479 GGTTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTTAGAAAGGAAACGAACTCCAAA 538  
Db 577 GGTTTTTTCTACAGATGAATTTGAAGAAATTTAAAGAACTTTAGAAAGGAAATGAACCTCCAAA 636  
Qy 539 GACATGAGATGAATTTCTTTTGGATTTAGACATCATGAAGGTCTATCATGACAGATC 598  
Db 637 GACATGAGATGAATTTCTTTTGGATTTAGGACATCATGAAGGTCTATATATGCGGATC 696  
Qy 599 TATACTACCTCAGTCACAGATGAGCAGTGAAGTGGCGGAAAGAAAGAACCCAAAGATC 658  
Db 697 TATACTACCTCAGTCACAGATGAGCAGTGAATTTGGCGGAAAGAAAGGCCAAAGATC 756  
Qy 659 TGACAGAGCTGGTCCAGCGGAGATAACATATCTGCAGAAATCCCAAGGACTGCGAGCAAG 718  
Db 757 TGACAGAGCTGGTTCAGCGGAGATACATATCTTCAGATCCCAAGGACTGCGAGCAAG 816  
Qy 719 CCAGAAAGCTGGTATGTAATTAATCAAAAGGCTGTGGCTATGGATGTCAACTCATCATG 778  
Db 817 CCAAAAGCTGGTGTGTAATTAATCAAAAGGCTGTGGCTATGGCTGTGAGCTTCCATCATG 876  
Qy 779 TGGTTTACTGCTTCATGATTTGCTTATGGCACCCAGCGAACACTCATCTTGAATCTCAGA 838  
Db 877 TGGTTTACTGCTTCATGATTTGCTTATGGCACCCAGCGAACACTCATCTTGAATCTCAGA 936  
Qy 839 ATTGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAGACCTGTAGTGAAGTGAATGTC 898  
Db 937 ATTGGCGCTATGCTACTGGGATGGGAGACTGTATTTAGGCGCTGTAGTGAAGTGAATGTC 996  
Qy 899 CAGACAGCTTGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTTC 958  
Db 997 CAGACAGCTTGGCTCTCCACTGGACACTGGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTTC 1056  
Qy 959 AAGTGGTGGAGTCCCATTTGTAGACAGCTCCATCTCGTCTCTTACTTACCTTGG 1018  
Db 1057 AAGTGGTGGAGTCCCATTTGTAGACAGCTCTTCATCCCGCTCTCCATATTTACCTTGG 1116  
Qy 1019 CTGTACCAAGAGACTTGCAGATCGACTCTCTGAGAGTCCATGGTGTCTCTGAGTGTGGT 1078  
Db 1117 CTGTACCAAGAGACTTGCAGATCGACTTGTACAGTGCATGGTGACCTCGAGTGTGGT 1176  
Qy 1079 GGGTATCCAGTTTGTCAAATACTTGTATCCGTCCACAACTTGGCTGAAAGGGAAATAG 1138  
Db 1177 GGGTGTCTCAGTTTGTCAAATACTTGTATCCGCCACACAGCTTGGCTAGAAAAGAAATAG 1236  
Qy 1139 AAGAAACCAACAGAACTTGGCTTCAAACATCAGTTTGGAGTCCATCTGACAGCA 1198  
Db 1237 AAGAAAGCCACAGAACTTGGCTTCAAACATCAGTTTGGAGTCCATCTGACAGCA 1296  
Qy 1199 CTGACAAAGTGGGAACAGAACAGCTTCCATCCATTTGAGGAAATACATGGTACAGTGTG 1258  
Db 1297 CAGACAAAGTGGGAACAGAACAGCTGCTTCCATCCATTTGAAGAGTACATGGTGCATGTG 1356  
Qy 1259 AAGAACATTTTTCAGCTTCTCGAACGCAAGTGAAGTGAATGAAGAGAGTGTATCTGG 1318  
Db 1357 AAGAACATTTTTCAGCTTCTTCACGCAAGTGAAGTGAAGTGAAGAGAGTGTATTTGG 1416  
Qy 1319 CCATGTATGACCTTCTTTTGTAAAGGAGGCAAGACAAAGTACTTCCAAATTTATGAATTA 1378  
Db 1417 CCACAGATGACCTTCTTTTAAAGGAGGCAAGAACAAAGTACTTCCAAATTTATGAATTA 1476

1379	TTAGTGTATAACTCTATTTCTTTGGTCACTGGA	CTACAAACCGATACACAGAAAATTCAC	1438
Qy			
1477	TTAGTGTATAACTCTATTTCTTGGTCACTGGA	CTGCAAAATCGATACACAGAAAATTCAC	1536
Db			
1439	TTCCGGGCGTGATCCCTGGATATACATTTCT	CTCCAGGCTGACTTCCTTGTGTGTAATT	1498
Qy			
1537	TTCCGTGGAGTGATCCTCGATATACATTTCT	CTCTCAGGAGACTTCCTAGTGTGTAATT	1596
Db			
1499	TTTTCATCCCAAGTCTGTGAGGTTGCTTTAT	GAAATCATGCAAAACACTGCATCTCTGATGCCT	1558
Qy			
1597	TTTTCATCCCAAGTCTGTGAGGTTGCTTTAT	GAAATCATGCAAAACACTGCATCTCTGATGCCT	1656
Db			
1559	CTGCAAACTTCCATTTCTTTAGATGACATCT	ACTACTATTTTGGAGGCCAAAATGCCCAACAAC	1618
Qy			
1657	CTGCAAACTTCCATTTCTTTAGATGACATCT	ACTACTATTTTGGAGGCCGAAATGCCCAACAATC	1716
Db			
1619	AGATTGCAGTTTATPCTTCACCAACTCGAA	CTAAAGAGAAAATCCCATGGAACCTTGGAG	1678
Qy			
1717	AAATTGCACTTTATGCTCACCACCCGGA	ACTGCAGATGAAATTTCCCATGGAACCTTGGAG	1776
Db			
1679	ATATCATTTGTTGGCTGGAAACCATTTGGA	ATGGTACTCTAAAGTGTTCAACAGAAAAC	1738
Qy			
1777	ATATCATTTGTTGGCTGGAAATCATTTGG	ATGGCTATTTCTAAAGTGTTCAACAGGAAAT	1836
Db			
1739	TAGAAAAACAGGCTGTACCTTCTCTACAA	AGTCGAGAGAGATAGAAAACAGTCAAAAT	1798
Qy			
1837	TGGGAGGACGGGCTTATATCCCTCTACA	AGTTCCAGAGAGATAGAAAACCGTCAAGT	1896
Db			
1799	ACCTCATATCTCTGAAGCTGAAAAATAG	AGATGGAGTGTAGAGATTAACAACAGAAAT	1858
Qy			
1897	ACCCACATATCTGTAGGCTGAGAAATAA	AGCTCACATGGAAGAGATAAACGCCAAACT	1956
Db			
1859	TAGTTTCAGACCATCTCAG	1876	
Qy			
1957	CAGTTCCGACCAACTCAG	1974	
Db			

## RESULT 5

```

US-09-839-136-1
; Sequence 1, Application US/09839136
; Patent No. US20020081694A1
; GENERAL INFORMATION:
; APPLICANT: Naoyuki TANIGUCHI et al.
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Pig
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1728)
US-09-839-136-1

```

[illegible]

Db 1141 GAATACACGGTGACGTTGAAGAAGACTTTTCAGTCTTCTGCTCCAGAAATGCAAGTGGAT 1200  
Qy 1300 AAAAAAGAGTGTATCTGCGCACTGATGACCCCTCTCTTTGTTAAAGAGGCAAGCAAAAG 1359  
Db 1201 AAAAAAGGGTATTTGGCCACAGATGACCTGCTTTGTTAAAGAGGCAAAAGCAAAAG 1260  
Qy 1360 TACTCCAATTATGAATTTATTAGTGATAACTCTATTTCTTGGTCAAGCTGGACTACACAAC 1419  
Db 1261 TACCCAGTTATGAATTTATTAGTGATAACTCTATCTCTTGGTCAAGCTGGACTACATAAT 1320  
Qy 1420 CGATACACAGAAATTCATCTCGGGCGTGATCCCTGGATATACACTTTCTCTCCAGGCT 1479  
Db 1321 CGATATACAGAAATTCATCTCGGGGTGTGATCCCTGGATATACACTTTCTCTCCAGGCA 1380  
Qy 1480 GACTTCTCTGTGTACTTTTTCATCCAGGCTCTAGGGTGTCTTATGAATCATGCAA 1539  
Db 1381 GACTTCTTAGTGTACTTTTTCATCGAGGCTGTAGAGTGTCTTATGAATCATGCAA 1440  
Qy 1540 ACATGTCATCTGATCGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGA 1599  
Db 1441 GCGTGTGATCTGATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTACTATTTTGA 1500  
Qy 1600 GGCAGAAATGCCCAACAGATTCAGTTTATCTCTACCAACCTCGAACTAAAGAGGAA 1659  
Db 1501 GGCAGAAATGCCCAACAAATTCGCCATTTATCTCTCAACACCTCGAACTGAAGGAGAA 1560  
Qy 1660 ATCCCATGGAACTGGAGATATCATTTGTTGGCTGGAAACCATTTGGAATGGTTACTCT 1719  
Db 1561 ATCCCATGGAACTGGAGATATCATTTGTTGGCTGGAAATCATCTGGATGGCTATCCT 1620  
Qy 1720 AAAGTGTCACAGAAAACTAGSAAAAACAGGCTGTACCCCTTCTACAAAAGTCCGAGAG 1779  
Db 1621 AAAGTGTTAAAGAAAACTGGAGGAGGCGGCTATATCTCTCTCAAAAGTTCGAGAG 1680  
Qy 1780 AAGTAGAAACAGTCAATACCTTACATATCTCTGAAGCTGAAATA 1826  
Db 1681 AAGTAGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAGTA 1727

RESULT 6

US-09-971-773-6  
; Sequence 6, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; PRIORITY FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIORITY FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIORITY FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 979  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-6

Query Match 48.6%; Score 976.4; DB 11; Length 979;  
Best Local Similarity 99.9%; Pred.No. 3.4e-273;  
Matches 977; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 819 ACTCATCTTGAATCTCAGAATTGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAG 878

Db 1 ACTCATCTTGAATCTCAGAATTGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAG 60  
Qy 879 ACCTGTAAGTCAGACATGACAGACAGGTCTGGGCTCTCCACTGACACTGGTCAAGTGA 938  
Db 61 ACCTGTAAGTCAGACATGACAGACAGGTCTGGGCTCTCCACTGACACTGGTCAAGTGA 120  
Qy 939 AGTGAAGACAAAAAATGTTCAAGTGGTGGAGCTCCCAATTTGTAGACAGCTCCATCTCG 998  
Db 121 AGTGAAGACAAAAAATGTTCAAGTGGTGGAGCTCCCAATTTGTAGACAGCTCCATCTCG 180  
Qy 999 TCCTCTCTTACCTTAATCCCTTGGCTGTACCAAGAAAGACCTTGCAGATCGACTCTCTGAGAGTCA 1058  
Db 181 TCCTCTCTTACCTTAATCCCTTGGCTGTACCAAGAAAGACCTTGCAGATCGACTCTCTGAGAGTCA 240  
Qy 1059 TGGTGTATCTCGAGTGGTGGGTATCCAGTGTGTCAAACTTGTCAAACTTGTCAAACTTGTCAAACT 1118  
Db 241 TGGTGTATCTCGAGTGGTGGGTATCCAGTGTGTCAAACTTGTCAAACTTGTCAAACTTGTCAAACT 300  
Qy 1119 TTGGCTGAAAGGAAATAGAAAGAAACCAACAAAGAGCTTGGCTTCAAACTTCAAACTTCAAACTTAT 1178  
Db 301 TTGGCTGAAAGGAAATAGAAAGAAACCAACAAAGAGCTTGGCTTCAAACTTCAAACTTCAAACTTAT 360  
Qy 1179 TGGAGTCCATGTCTAGACGCACTGACAAAGTGGGAAAGAGAGGCTTCCATCCCATTTGA 1238  
Db 361 TGGAGTCCATGTCTAGACGCACTGACAAAGTGGGAAAGAGAGGCTTCCATCCCATTTGA 420  
Qy 1239 GGAATACATGTGTAACGTTGAAGAACATTTTTCAGTCTTCGAAACGCAAGTGAAGTGA 1298  
Db 421 GGAATACATGTGTAACGTTGAAGAACATTTTTCAGTCTTCGAAACGCAAGTGAAGTGA 480  
Qy 1299 TAAAAAAGAGTGTATCTGGCCACTGTAGTACCTTCTTTGTTTAAAGAGGCAAAAGACAAA 1358  
Db 481 TAAAAAAGAGTGTATCTGGCCACTGTAGTACCTTCTTTGTTTAAAGAGGCAAAAGACAAA 540  
Qy 1359 GTACTCCAAATTAATGAATTTATTAGTGATACTTATTTCTTGGTCAAGTGAAGTGA 1418  
Db 541 GTACTCCAAATTAATGAATTTATTAGTGATACTTATTTCTTGGTCAAGTGAAGTGA 600  
Qy 1419 CCGATACACAGAAATTCACCTTCGGGGGTGATCTCGATATACACTTTCTCTCCAGGC 1478  
Db 601 CCGATACACAGAAATTCACCTTCGGGGGTGATCTCGATATACACTTTCTCTCCAGGC 660  
Qy 1479 TGACTTCTCTGTGTACTTTTTCATCCAGGTCTGTAGGGTGTCTTATGAATCATGCA 1538  
Db 661 TGACTTCTCTGTGTACTTTTTCATCCAGGTCTGTAGGGTGTCTTATGAATCATGCA 720  
Qy 1539 AACACTGCATCTCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 1598  
Db 721 AACACTGCATCTCTGATGCTCTGCAAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 780  
Qy 1599 AGGCAAAATGCCACAAACAGATGCTGAGTTTATCTCCACCAACCTCGAACTAAAGAGA 1658  
Db 781 AGGCAAAATGCCACAAACAGATGCTGAGTTTATCTCTCACCACCTCGAACTAAAGAGA 840  
Qy 1659 AATCCCATGGAACCTTGAGATATCATTTGGTGTGGCTGGAACCATTTGGAATGGTTACTC 1718  
Db 841 AATCCCATGGAACCTTGAGATATCATTTGGTGTGGCTGGAACCATTTGGAATGGTTACTC 900  
Qy 1719 TAAAGGTGTCAACAGAAATCTAGGAAACAGGCTGTACCCCTTCTCTCAAAAGTCCGAGA 1778  
Db 901 TAAAGGTGTCAACAGAAATCTAGGAAACAGGCTGTACCCCTTCTCTCAAAAGTCCGAGA 960  
Qy 1779 GAAGATGAACAGTCA 1796  
Db 961 GAAGATGAACAGTCA 978

RESULT 7  
US-09-971-773-7  
; Sequence 7, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:

```
/ APPLICANT: Yutaka KANDA
/ APPLICANT: Mitsuo SATOH
/ APPLICANT: Kazuyasu NAKAMURA
/ APPLICANT: Kazuhisa UCHIDA
/ APPLICANT: Toyohide SHINKAWA
/ APPLICANT: Naoko YAMANE
/ APPLICANT: Motoko YAMASAKI
/ APPLICANT: No. US20030115614ALUO HANAI
/ TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
/ FILE REFERENCE: 249-202
/ CURRENT APPLICATION NUMBER: US/09/971,773
/ CURRENT FILING DATE: 2002-08-30
/ PRIOR APPLICATION NUMBER: JP 2000-308526
/ PRIOR FILING DATE: 2000-10-06
/ PRIOR APPLICATION NUMBER: US 60/268,926
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 73
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 7
/ LENGTH: 979
/ TYPE: DNA
/ ORGANISM: Rattus norvegicus
US-09-971-773-7

Query Match      44.5%; Score 893.2; DB 11; Length 979;
Best Local Similarity 94.6%; Pred. No. 5.8e-249;
Matches 925; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 819 ACTCATCTTGGAAATCTCAGAAATGGCGTATGCTACTGAGGATGGGAGACTGTTTAG 878
DB 1 ACTCATCTTGGAAATCTCAGAAATGGCGTATGCTACTGAGGATGGGAGACTGTTTAG 60
QY 879 ACCTGTAAGTGAAGATGACACAGACAGAGTCTGGCTCTCCACTGACACTGTCAGGTGA 938
DB 61 ACCTGTAAGTGAAGATGACACAGACAGATCTGGCTCTCCACTGACACTGTCAGGTGA 120
QY 939 AGTGAAGGACAAAATATGTTCAAGTGGTGGAGTCCCGATTTGATGACAGCCCTCCATPCTCG 998
DB 121 AGTGAATGACAAAATATTTCAAGTGGTGGAGTCCCGATTTGATGACAGCCCTCCATPCTCG 180
QY 999 TCCTCTTACTTACCTTGGCTGTACACAGAGACCTTGCAGATCGACTCTCGAGAGTCCA 1058
DB 181 GCCTCTTACTTACCACTGGCTGTTCAGAGAGACCTTGCAGATCGACTCGTAAGAGTCCA 240
QY 1059 TGGTGCATCTGCAGTGTGGTGGTATCCAGTTCCTGCTCAAAATCTTGATCCGTCACAAACC 1118
DB 241 TGGTGCATCTGCAGTGTGGTGGTGTCCAGTTCCTGCTCAAAATCTTGATTCGTCACAAACC 300
QY 1119 TTGGCTGGAAGGGAATAGAAAGAAACCAAGAAAGCTTGGCTTCAAAACATCCAGTTAT 1178
DB 301 TTGGCTAGAAAGGAATAGAAAGAAAGCCACCAAGAAAGCTTGGCTTCAAAACATCCAGTCA 360
QY 1179 TGGAGTCCATGTACAGCAGCTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATGA 1238
DB 361 TGGAGTCCATGTACAGCAGCTGACAAAGTGGGAACAGAGCAGCTTCCATCCCATGA 420
QY 1239 GGAATACATGTACAGTGTGAAGAACATTTTCAGCTTCTCGAAACGCAAGATGAAGTGA 1298
DB 421 AGAGTACATGGTACATGTTGAAGAACATTTTCAGCTTCTCGAACGCAAGATGAAGTGA 480
QY 1299 TAAAAAAGAGGTATCTGGCCACTGATGACCCCTTCTTTGTTTAAAGAGGCAAGACAAA 1358
DB 481 TAAAAAAGAGGTATCTGGCTACCGATGACCCCTTCTTTGTTTAAAGAGGCAAGACAAA 540
QY 1359 GTACTCCAAATATGATTTATTTAGTGATTAATCTATTTCTGGTTCAGCTGACTACAAA 1418
DB 541 GTACTCCAAATATGATTTATTTAGTGATTAATCTATTTCTGGTTCAGCTGACTACAAA 600
QY 1419 CGGATACACAGAAAATTCACCTCGGGGCTGATCTCGGATATACACTTCTCTCCAGGC 1478
DB 601 TCGGTACACAGAAAATTCACCTCGGGGCTGATCTCGGATATACACTTCTCTCCAGGC 660
QY 1479 TGACTTCCTTGTGTGATCTTTTTCATCCAGGCTGTGTAGGGTGTCTTATGAATCATGCA 1538
```

```
DB 661 TGACTTCTAGTGTGTACTTTTTCATCCAGGCTGTGCGGTTGCTTTATGAATCATGCA 720
QY 1539 AACACTGCATCTGATGCTCTGCAAACTTCCTATCTTTTAGATGACATCTACTATTTGG 1598
DB 721 AACCTGATCTGATGCTCTGCAAACTTCCTATCTTTTAGATGACATCTACTATTTGG 780
QY 1599 AGGCCAAAATGCCCAACACAGATTGCGATTATCTCTCACCAACTCGAACTAAAGAGA 1658
DB 781 AGGCCAAAATGCCCAACACAGATTGCGATTATCTCTCACCAACTCGAACTGATGAGA 840
QY 1659 AATCCCATGGAACCTGGAGATATCATTTGCTGGCTGGAAACCATTTGGAATGTTACTC 1718
DB 841 AATTCCAATGGAACTGGAGATATCATTTGCTGGCTGGAAACCATTTGGAATGTTACTC 900
QY 1719 TAAAGGTGTCAACAGAAAACCTAGCAAAAACAGGCTGTACCCCTTCTACAAAGTCCGAGA 1778
DB 901 TAAAGGTGTCAACAGAAAACCTTGGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGA 960
QY 1779 GAAGATAGAAACAGTCAA 1796
DB 961 GAAGATAGAAACGGTCAA 978

RESULT 8
US-09-918-995-33468
; Sequence 33468, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33468
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33468
```

```
Query Match      18.3%; Score 368.2; DB 11; Length 422;
Best Local Similarity 92.2%; Pred. No. 2.5e-96;
Matches 388; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 863 GGGAGACTGTGTTTAGACCTGTAAAGTGAGACATGCACAGACAGTCTGGCCTCTCCACTG 922
DB 2 GGGAGACTGTATTTTAGGCTGTAAAGTGAGACATGCACAGACAGATCTGGCATCTCCACTG 61
QY 923 GACACTGTCAGGTGAAGTGAAGCAAAAATGTTCAAGTGGTTCAGCTCCGCTCCCATTTAG 982
DB 62 GACACTGTCAGGTGAAGTGAAGCAAAAATGTTCAAGTGGTTCAGCTCCGCTCCCATTTAG 121
QY 983 ACAGCTTCATCTCTGCTCTCTTACTTTACCTTTGGCTGTACCAAGAGACCTTGCAGATC 1042
DB 122 ACAGTCTTCATCTCCGCTCTCTCCATATTTACCTTTGGCTGTACCAAGAGACCTTGCAGATC 181
QY 1043 GACTCTCTGAGAGTCCATGGTGATCTCTGAGTGTGGGTATCCAGTTTCTCAAAATCT 1102
DB 182 GACTTGTACGAGTGCATGGTGACCTCGAGTGTGGTGTCTCAGTTTCTCAAAATCT 241
QY 1103 TGATCCGTCCACCAACCTTGGCTGGAAGGGAATAGAGAAACCAACCAAGAGCTTGGCT 1162
DB 242 TGATCCGCCCAACAGCTTGGCTAGAAAAGAAAATAGAAAGAGCCCAAGAGCTTGGCT 301
QY 1163 TCAAAACATCCAGTTATTTGGAGTCCATGTTCAGACGCACTGCACAAAAGTGGAAACAGAAAGCAG 1222
DB 302 TCAAAACATCCAGTTATTTGGAGTCCATGTTCAGACGCACTGCAGAAAAGTGGAAACAGAAAGCTG 361
```





Db 186 TTGGTGGCTGGAATCATTCGGATGGCTATTCTAAAGGTGTCAACAGGAAATTGGGA 245  
Qy 1745 AAACAGGCTGTACCTCTTCAAGTCCGAGAGAGATAGAAACAGTCAAAATACCCTA 1804  
Db 246 GGACGGGCTATATCTCTTCAAGTTCGAGAGAGATAGAAACGGTCAAGTACCCCA 305  
Qy 1805 CATATCTCTGAAGTCAAAATAGAGATGGAGTGTAAAGAGATTAAACACAGAAATTTAGTTC 1864  
Db 306 CATATCTCTGAAGTCAAAATAGAGATGGAGTGTAAAGAGATTAAACACAGAAATTTAGTTC 1864  
Qy 1865 AGACCATCTCAG 1876  
Db 366 GACCAAACTCAG 377

RESULT 11  
US-09-864-761-13292  
; Sequence 13292, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 13292  
; LENGTH: 551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

; OTHER INFORMATION: MAP TO AL109847.3  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
US-09-864-761-13292  
Query Match 10.9%; Score 218.8; DB 9; Length 551;  
Best Local Similarity 81.6%; Pred. No. 9.5e-53;  
Matches 253; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
Qy 872 TGTTTAGACCTGTAAAGTGAGACATGCACAGACAGGTCTGGCCTCTCCACTGGACACTGGT 931  
Db 116 TGTGAATGGTGATGTATAGGAATACCATGTGTATATATATATATATATATATATAT 175  
Qy 932 CAGGTGAAGTGAAGGACAAAATTTCAAGTGTGAGTCCCAATTTGATAGACAGCTCC 991  
Db 176 CAGGTGAAGTGAAGGACAAAATTTCAAGTGTGAGTCCCAATTTGATAGACAGCTCC 235  
Qy 992 ATCTCTGCTCTCTTACTTACCCTTGGCTGTACAGAGACCTTGCAGATCGACTCTCTGA 1051  
Db 236 ATCCCTGCTCTCCATATTTACCCTTGGCTGTACAGAGACCTTGCAGATCGACTTGTAC 295  
Qy 1052 GAGTCCATGGTGATCCTGCAAGTGTGGGTATCCCAAGTTTGTCAAATACTTGTATCGGT 1111  
Db 296 GAGTCCATGGTGATCCTGCAAGTGTGGGTATCCCAAGTTTGTCAAATACTTGTATCGGT 355  
Qy 1112 CACAACCTTGGCTGGAAGGAAATAGAGAAACCAAGAAAGCTTGGCTTCAAAATC 1171  
Db 356 CACAGCCTTGGCTAGAAAAGAAATAGAGAAAGCCCAAGAGCTTGGCTTCAAAATC 415  
Qy 1172 CAGTTATTGG 1181  
Db 416 CAGTTATTGG 425

RESULT 12  
US-09-864-761-18181  
; Sequence 18181, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 18181  
; LENGTH: 551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:



```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18181
; LENGTH: 248
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUATE 6.50e+00
; OTHER INFORMATION: NT HIT: AF038280.1, EVALUATE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW387766.1, EVALUATE 0.00e+00
US-09-864-761-18181

Query Match 10.6%; Score 213.4; DB 9; Length 248;
Best Local Similarity 91.5%; Pred. No. 2.1e-51;
Matches 226; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 935 GTGAAGTGAAGACAAAATGTTCAAGTGTGCGAGCTCCCATTTGTAGACAGCTCCATC 994
DB 1 GTGAAGTGAAGACAAAATGTTCAAGTGTGCGAGCTCCCATTTGTAGACAGCTCTTATC 60

QY 995 CTCGTCTCTTACTTACCCTTGGCTGTACCAAGAGACTTGCAGATCGACTCTTGAGAG 1054
DB 61 CCCGTCCTCCATATTTACCCTTGGCTGTACCAAGAGACTTGCAGATCGACTTGTACGAG 120

QY 1055 TCCATGGTGATCCTGCAGTGTGGGTATCCAGTTTGTCAAATCTTGATTCGCTCCAC 1114
DB 121 TGCATGGTGACCTGCAGTGTGGGTGTCTCAGTTTGTCAAATCTTGATTCGCTCCAC 180

QY 1115 AACCTTGGCTGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAACATCCAG 1174
DB 181 AGCCTTGGCTAGAAAAGAAATAGAGAAACCAAGAGCTTGGCTTCAACATCCAG 240

QY 1175 TTATTGG 1181
DB 241 TTATTGG 247

RESULT 13
US-09-864-761-30284
; Sequence 30284, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomicca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30284
; LENGTH: 248
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUATE 6.50e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW387766.1, EVALUATE 0.00e+00
; OTHER INFORMATION: NT HIT: AF038280.1, EVALUATE 0.00e+00
US-09-864-761-30284

Query Match 10.6%; Score 213.4; DB 9; Length 248;
Best Local Similarity 91.5%; Pred. No. 2.1e-51;
Matches 226; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 935 GTGAAGTGAAGACAAAATGTTCAAGTGTGCGAGCTCCCATTTGTAGACAGCTCCATC 994
DB 1 GTGAAGTGAAGACAAAATGTTCAAGTGTGCGAGCTCCCATTTGTAGACAGCTCTTATC 60

QY 995 CTCGTCTCTTACTTACCCTTGGCTGTACCAAGAGACTTGCAGATCGACTCTTGAGAG 1054
DB 61 CCCGTCCTCCATATTTACCCTTGGCTGTACCAAGAGACTTGCAGATCGACTTGTACGAG 120

QY 1055 TCCATGGTGATCCTGCAGTGTGGGTATCCAGTTTGTCAAATCTTGATTCGCTCCAC 1114
DB 121 TGCATGGTGACCTGCAGTGTGGGTGTCTCAGTTTGTCAAATCTTGATTCGCTCCAC 180

QY 1115 AACCTTGGCTGAAGGGAATAGAGAAACCAAGAGCTTGGCTTCAACATCCAG 1174
DB 181 AGCCTTGGCTAGAAAAGAAATAGAGAAACCAAGAGCTTGGCTTCAACATCCAG 240

QY 1175 TTATTGG 1181
DB 241 TTATTGG 247
```

```
Db      241 TTATTGG 247

RESULT 14
US-10-125-968-274
; Sequence 274, Application US/10125968
; Publication NO. US20030215805A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Palermo, Adam
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; APPLICANT: Elias, Josh
; APPLICANT: Mertens, Maureen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-032
; CURRENT APPLICATION NUMBER: US/10/125,968
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,163
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 1417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204, 266, 320, 328, 402, 424, 462, 470
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-274

Query Match      7.4%; Score 149; DB 13; Length 481;
Best Local Similarity 84.9%; Pred. No. 1.8e-32;
Matches 213; Conservative 0; Mismatches 32; Indels 6; Gaps 4;

Qy      338 CAGGAAGAGTCGCTGTTTAGAAGAACACAGCTGTTTAAAGCCAAAGACAG-ATTGAAAAT 396
Db      19  CGGCGAGGTACGGTTTTACAAGAGAGCAGCTGTTTAAAGCCAAAGACAGATTGAAAAT 78

Qy      397 TACAAGAAACAGCTAGG--AATGATCTGGGAAAGATCATGAATCTTAAAGGAGGAGGA 454
Db      79  TACAAGAAACAGACAGCAGTAATGCTCTGGGAAGGATCATGAATCTTAAAGGAGGAGGA 138

Qy      455 TTGAATAAGGAGCTAAAGAGCTCTGTTTTTCTACAAGTCAATTAAGAAATTAAGA 514
Db      139 TTGAATAAGGAGCTAAAGAGCTCTGTTTTTCTACAAGTCAATTAAGAAATTAAGA 198

Qy      515 AATTAGAAGGAACGAACCTCCAAAGACATGC-AGATGAATCTTTTGG--ATTTAGGAC 571
Db      199 ACTTANAGGAATGACTCCAAAGACATGCAAGATGAATTTCTTTTGGATTTAGGAC 258

Qy      572 ATCATCAAGG 582
Db      259 ATCATGANAAG 269

RESULT 15
US-09-864-761-1421
; Sequence 1421, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1421
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
US-09-864-761-1421

Query Match      6.8%; Score 136.4; DB 9; Length 384;
Best Local Similarity 76.6%; Pred. No. 7.4e-29;
Matches 167; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy      872 TGTTTAGACCTGTAAGTGAGACATGCACAGACAGCTGGCCTCTCCACTGGACACTGGT 931
Db      167 TGTTGAATGGTGGATGTATAGGAATACCATGTGTATATATATATATATTTCAATATTTGT 226

Qy      932 CAGGTGAAGTGAAGCAACAAAATGTTCAAGTGGTCGAGCTCCCATTTGTAGACAGCTCC 991
Db      227 CAGGTGAAGTGAAGCAACAAAATGTTCAAGTGGTCGAGCTTCCTCATTTGTAGACAGTCTTC 286

Qy      992 ATCTCGTCTCTCTTACTTACCTTGGCTGTACAGAGACCTTTCAGATCGACTCTGTA 1051
Db      287 ATCCCGTCTCTCCATATTTACCTTGGCTGTACAGAGACCTTCGAGATCGACTTGTAC 346

Qy      1052 GAGTCCATGGTGTATCTTCGAGTGTGGTGGGTATCCCG 1089
```

Db 347 GAGTGCATGGTGACCCCTGCAGTGGTGGTGCTCTCAG 384

Search completed: February 2, 2004, 19:02:53  
Job time : 717.477 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM\_nucleic - nucleic search, using sw model

Run on: February 2, 2004, 11:54:20 ; Search time 4312.69 Seconds  
(without alignments)  
11316.242 Million cell updates/sec

Title: US-09-971-773-1  
Perfect score: 2008  
Sequence: 1 aacagaactatttctg.....gctgtgcctcaagcccatg 2008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_esta:\*  
2: em\_estum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1809.4	90.1	2766	11 AK051811	AK051811 Mus muscu
2	1809.4	90.1	3052	11 AK048520	AK048520 Mus muscu
3	831.6	41.4	937	13 B0892101	B0892101 AGENCOURT
4	810	40.3	1066	12 B928521	B928521 AGENCOURT

5	768	38.2	940	13	BQ653329	BQ653329 AGENCOURT
6	757.6	37.7	900	13	BX452321	BX452321 AGENCOURT
7	751.2	37.4	886	13	BQ423860	BQ423860 AGENCOURT
8	741	36.9	923	13	BX429921	BX429921 AGENCOURT
9	741	36.9	1027	13	BX450010	BX450010 AGENCOURT
10	738	36.8	855	13	BUI184296	BUI184296 AGENCOURT
11	730.2	36.4	898	14	CB194828	CB194828 AGENCOURT
12	699.2	34.3	834	9	AUI124128	AUI124128 AGENCOURT
13	641.2	31.9	937	13	BUI13162	BUI13162 AGENCOURT
14	634.6	31.6	757	13	BX414472	BX414472 AGENCOURT
15	634.6	31.6	790	14	CB989769	CB989769 AGENCOURT
16	632	31.5	752	13	BQ604588	BQ604588 MI-P-CP1-
17	625.4	31.1	965	12	B1546364	B1546364 60318856
18	623.2	31.0	692	14	CB105611	CB105611 K-EST0130
19	623.2	31.0	692	14	CB105621	CB105621 K-EST0130
20	612	30.5	717	14	CB322023	CB322023 UI-CF-PNO
21	604.6	30.1	970	13	BQ926824	BQ926824 AGENCOURT
22	601.6	30.0	656	13	BQ552954	BQ552954 H4019310-
23	599.4	29.9	702	12	BG973221	BG973221 602842563
24	593.6	29.6	805	14	CA489662	CA489662 AGENCOURT
25	592.2	29.0	936	10	BE893107	BE893107 601437002
26	573.6	28.6	769	10	CB105920	CB105920 K-EST0135
27	572.2	28.7	645	14	CB105920	CB105920 K-EST0135
28	567.8	28.2	731	13	BX456265	BX456265 BX456265
29	566.6	28.2	637	14	CB105617	CB105617 K-EST0130
30	563.4	28.1	613	14	CA539475	CA539475 C0279008-
31	559.6	27.9	622	14	CB105602	CB105602 K-EST0130
32	559	27.8	715	12	BG771085	BG771085 602719363
33	558.6	27.8	621	14	CB105873	CB105873 K-EST0135
34	555.6	27.7	618	14	CB105618	CB105618 K-EST0130
35	551	27.4	938	13	BX433236	BX433236 BX433236
36	548.8	27.3	608	14	CB105918	CB105918 K-EST0135
37	547.8	27.3	607	14	CB105608	CB105608 K-EST0130
38	547.8	27.3	607	14	CB105882	CB105882 K-EST0135
39	547.6	27.2	900	13	BQ731540	BQ731540 AGENCOURT
40	546.2	27.2	607	14	CB105899	CB105899 K-EST0135
41	544.8	27.1	653	10	BG619839	BG619839 602618014
42	540	26.9	610	14	CB105867	CB105867 K-EST0135
43	528.4	26.3	586	14	CB105886	CB105886 K-EST0135
44	526	26.2	585	14	CB105871	CB105871 K-EST0135
45	524	26.1	573	9	AI122135	AI122135 udl5d10.r

## ALIGNMENTS

### RESULT 1

AK051811

LOCUS

DEFINITION

AK051811 2766 bp mRNA linear HTC 05-DEC-2002  
Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length  
enriched library, clone:D330003K17 product:fucoyltransferase 8,  
full insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AK051811.1 GI:26342229  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

REFERENCE  
AUTHORS

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Iehi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,  
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,  
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,  
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,  
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,  
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,  
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,  
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,  
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,  
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,  
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,  
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Buit,C.,  
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,  
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,  
Marchionni,L., Mashima,J., Mazzarelli,J., Mombauts,P., Nordone,P.,  
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,  
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,  
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,D.,  
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.  
and Hayashizaki,Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2766)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitho,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submision  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

## FEATURES

## source

1. 2766  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM,DB:D230003K17"  
/db\_xref="taxon:10090"  
/clone="D230003K17"  
/tissue\_type="eyeball"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
473 - 2200  
/note="unnamed protein product; fucosyltransferase 8  
(MGDI:1858901, GB|NM\_016893, evidence: BLASTN, 99%,  
match=2175)  
putative"  
/codon\_start=1  
/protein\_id="BAC34777.1"  
/db\_xref="GI:26342230"  
/translation="MRAATGSGWIMILFAMGTLFLLYIGHLVRDNDHPDHSRELS  
KILAKLERKQNEEDLRMAESLRIPGPIQGTATGRVLEQLVAKKEIQEINYKK  
QARNGLKDEHILRRRIENGAKELWFFLOSELKKLEHLEQLORHADELILLGHH  
RSIMTLYLSQTDGAGDREKAKOLTELVRRIYQLONPKDCSKARKLVNKNKGC  
GYGCOLHHVYCFMIAVGTORTLILESONRYATGWTETFRPVSECTDRSGLSGTH  
WSSEVNDKNQVVELPIVDSLHPRPPYPLPLAVPEDLADRLLRVHGDPAVWVSQFVKY  
LIRPQWLEKEIEEATKKGFKHPVIGVHRVRTDKVGEARAFHFIETWVHEHFQL  
LARMQVKRVYLATDDPTLLKEATKYSNYSFISNLSWSAGLHNRITENSLRGV  
ILDTHIQADFLVCTPFSQVCRVAYEIMQTLHPDASANFHSLLDIYFGQNAHQI  
AVYHPKRTETETPEPEDIIGVAGNHWGYSKGNRKLKGTGLYPSYKVKREKLETVK  
YPTYPEAK"  
2750 . 2755  
/note="putative"  
2766

## CDS

BASE COUNT 838 a 591 c 619 g 718 t  
ORIGIN  
polyA\_signal  
polyA\_site  
Query Match 90.1%; Score 1809.4; DB 11; Length 2766;  
Best Local Similarity 94.1%; Pred. No. 0;  
Matches 1891; Conservative 0; Mismatches 116; Indels 2; Gaps 1;  
Qy 1 AACAGAACTATTCTCTGTGGCTAACTAGAACACAGAGTACAAGTTTCCAACTCTT 60  
Db 372 AAAAGAAACCTATTCTTCTGTGGCTAACTAGAACACAGAGTACAAGTTTCCAGTCTT 431  
Qy 61 TGAGCTCCGAGAGACACAGA--AGGAGTTGAAACTCTGAAATCGGGGCATGGACTGTT 118  
Db 432 TGAGCTCCGAGAGATAGAGACAGAGTTGAAACTCTGAAATCGGGGCATGGACTGTT 491  
Qy 119 CCTGGCGTTGGATTATGCTCATTTCTTTTCTGGGGGACCTTATTTGTTTATATAGTG 178  
Db 492 CCTGGCGTTGGATTATGCTCATTTCTTTTCTGGGGGACCTTATTTTATATAGTG 551  
Qy 179 GTCAATTTGTTTCAGATAATGACACACCTGACCATTTCTAGCAGAGAACTCTCCAAGATTC 238  
Db 552 GTCAATTTGTTTCAGATAATGACACACCTGACCATTTCTAGCAGAGAACTCTCCAAGATTC 611  
Qy 239 TTGCAAGCTGGAGCGCTTAAACCAACAAATGAAGACTTGAAGAGAAATGCTGAGTCTC 298  
Db 612 TTGCAAGCTTGAACGCTTAAACACAGCAAAATGAAGACTTGAAGCGAATGGCTGAGTCTC 671  
Qy 299 TCCGAATACCAAGAGGCCCTTATTGATCAGGGGACAGCTACAGGAAGAGTCCGCTGTTTAG 358  
Db 672 TCCGAATACCAAGAGGCCCTTATTGATCAGGGGACAGCTACAGGAAGAGTCCGCTGTTTAG 731  
Qy 359 AAGAACAGCTGTTTAAAGCCCAAGAACAGATTGAAATTTACAAGAAACAAGCTAGGAATG 418  
Db 732 AAGAACAGCTGTTTAAAGCCCAAGAACAGATTGAAATTTACAAGAAACAAGCTAGGAATG 791  
Qy 419 ATCTGGAAAGGATCATGAAATCTTAAGGAGGAGGATTGAAATTTGAGAGCTCTTAAGAGCTCT 478  
Db 792 GTCTGGGAGAGGATCATGAAATCTTAAGAGGAGGATTGAAATTTGAGAGCTCTTAAGAGCTCT 851  
Qy 479 GGTTTTTCTACAAAGTGAATTTGAAGAAATTAAGAAATTTAGAAAGAAACGAACTCCAA 538

Db 852 GGTATTTCTACAAAGCGAACTGAAGAAATTAAGACCAATTAAGAGCAATGAATCTCCAAA 911  
Qy 539 GACATGCAGATGAATCTTTTGGATTAGACATCATGAAGTCTATCATGACATC 598  
Db 912 GACATGCAGATGAATCTTTTGGATTAGACATCATGAAGTCTATCATGACATC 971  
Qy 599 TATACACTCAGTCAACAGATGAGCAGGTGAGTGGCGGGAAGAAAGGCCAAAGATC 658  
Db 972 TATACTACTCAGTCAACAGATGAGCAGGTGAGTGGCGTGAAGAGGCCAAAGATC 1031  
Qy 659 TGACAGAGTGGTCCAGCGGGAATTAATATCTGCAGAAATCCCAAGACCTGCAGCAAG 718  
Db 1032 TGACAGAGTGGTCCAGCGGGAATTAATATCTGCAGAAATCCCAAGACCTGCAGCAAG 1091  
Qy 719 CCAGAAAGTGGTATGTAATCAACAAGGCTGTGGCTATGGATGTCATCTCCATCATG 778  
Db 1092 CCAGAAAGTGGTATGTAATCAACAAGGCTGTGGCTATGGATGTCATCTCCATCATG 1151  
Qy 779 TGGTTACTGCTTCATGATGCTTATGGCACCAGCGAACTCATCTTTGGAATCTCAGA 838  
Db 1152 TGGTCTACTGTTTATGATGCTTATGGCACCAGCGAACTCATCTTTGGAATCTCAGA 1211  
Qy 839 ATTGCGCTATGCTACTGAGATGGGAGACTGTGTTTGAACCTGTAAGTGAGACATGCA 898  
Db 1212 ATTGCGCTATGCTACTGAGATGGGAGACTGTGTTTGAACCTGTAAGTGAGACATGTA 1271  
Qy 899 CAGACAGTCTGGCTCTCCACTGGACACTGGTCAGGTGAAGTGAAGACAAAATGTC 958  
Db 1272 CAGACAGTCTGGCTCTCCACTGGACACTGGTCAGGTGAAGTGAAGACAAAATGTC 1331  
Qy 959 AAGTGGTCCAGCTCCCTCCCTTGTAGACAGCTCCATCTCTCTCTTACTTACCTCTGG 1018  
Db 1332 AAGTGGTCCAGCTCCCTCCCTTGTAGACAGCTCCATCTCTCTCTTACTTACCTCTGG 1391  
Qy 1019 CTGTACCAAGAGACCTTCAGATCGACTCTCTGAGAGTCCATGGTGAATCCTCGAGTGGT 1078  
Db 1392 CTGTTCAGAGAGACCTTCAGACAGCTCTCTAAGAGTCCATGGTGAACCTCGAGTGGT 1451  
Qy 1079 GGGTATCCAGTTTGTCAAACTACTGTATCGTCCACACCTTGGCTGGAAGGGAATAG 1138  
Db 1452 GGGTATCCAGTTTGTCAAACTACTGTATCGTCCACACCTTGGCTGGAAGGGAATAG 1511  
Qy 1139 AAGAAACCAACAAAGAGCTGGCTTCAAACTATGAGTATGGAGTCCATGTCCAGACCA 1198  
Db 1512 AAGAAACCAACAAAGAGCTGGCTTCAAACTATGAGTATGGAGTCCATGTCCAGACCA 1571  
Qy 1199 CTGACAAAGTGGGAACAGAGCAGCTTCCATCCATTCAGGAAATACATGGTACACGTTG 1258  
Db 1572 CAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGAGGAGTACATGGTACACGTTG 1631  
Qy 1259 AAGAACTTTTTCAGCTTCTCGAACCGCAAGTGAAGTGAATGAAGAGTGTATCTGG 1318  
Db 1632 AAGAACTTTTTCAGCTTCTCGAACCGCAAGTGAAGTGAATGAAGAGTGTATCTGG 1691  
Qy 1319 CCATGTAGCCCTCTTTTGTAAAGAGGAGCAAGACAAAGTACTCCCAATTTATGAATTA 1378  
Db 1692 CTACTGATGATCTTCTTTTGTAAAGAGGAGCAAGACAAAGTACTCCCAATTTATGAATTA 1751  
Qy 1379 TTAGTGAATACTTATTTTCTGGTCAGCTGGACTACACACCGATACACAGAAATTCAC 1438  
Db 1752 TTAGTGAATACTTATTTTCTGGTCAGCTGGACTACACACCGTACACAGAAATTCAC 1811  
Qy 1439 TTGGGGGCGTATCTCGATATACACTTTCTCTCCAGGCTGACTTCTTGTGTACTT 1498  
Db 1812 TTGGGGGCGTATCTCGATATACACTTTCTCTCAGAGCTGACTTTCTAGTGTGTACTT 1871  
Qy 1499 TTTCAATCCAGTCTGAGGTGCTTATGAATCATGCAACACTGATCTCTGATGCT 1558  
Db 1872 TTTCAATCCAGTCTGAGGTGCTTATGAATCATGCAACACTGATCTCTGATGCT 1931  
Qy 1559 CTGCAAACTCTCATTTCTTAGATGACATCTACTATTTTGGAGGCCAAATGCCCCAAC 1618

Db 1932 CTGCGAACTTTCCATTTCTTTGGATGACATCTACTATTTTGGAGGCCAAATGCCCAATC 1991  
Qy 1619 AGATTGCAGTTTATCTCTCACCAACTGAACTAAAGAGGAAATCCCAATGGAACCTGGAG 1678  
Db 1992 AGATTGCAGTTTATCTCTCACCAACTGAACTGAAAGGAAATTTCCATGGAACCTGGAG 2051  
Qy 1679 ATATCATTTGGTGTGGCTGGAAACCATTTGGAATGTTTACTCTTAAAGGTCTCAACAGAAAC 1738  
Db 2052 ATATCATTTGGTGTGGCTGGAAACCATTTGGGATGTTTATCTTAAAGGTATCAACAGAAAC 2111  
Qy 1739 TAGGAAAACAGGCGCTGTACCTTCTACAAAGTCCGAGAGAGATAGAAACAGTCAAT 1798  
Db 2112 TTGAAAACAGGCGCTGTATATCTCTTCTACAAAGTCCGAGAGAGATAGAAACAGTCAAT 2171  
Qy 1799 ACCCTACATATCTCTGAAAGTGAAGATAGAGATGGAGTGAAGATTAACAAACAGAAAT 1858  
Db 2172 ATCCACATATCTCTGAAAGTGAAGATAGAGATGAAGTGAAGATTAACAAACAGAAAT 2231  
Qy 1859 TAGTTTCAGACCATCTTAGCCAAAGCAGAGAGCCAGACTTAACATATGTTTTCATTCAGACAG 1918  
Db 2232 CACTTCAGACCATCTTAGCCAAAGCAGAGAGCCAGACTTAACATGTTTTCATTCAGACAG 2291  
Qy 1919 ATGCTCCGACCAAGAGCAAGTGGGAACCTTCAGATGCTGCACTGGTGGAGCGCTCTTT 1978  
Db 2292 ACGTCCACACAGAGCAAGCGGGAACCTTCAGATGCTGCACTGGTGGAGCGCTCTTT 2351  
Qy 1979 GTGAAGGCGTCTCTGCTCCCTCAAGCCCAT 2007  
Db 2352 ATGAAGGCGTGTGGTTCCTCAAGCCCAT 2380

RESULT 2  
AK048520  
LOCUS  
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:Cl:0068P22 product:fucoyltransferase 8, full insert sequence.  
ACCESSION AK048520  
VERSION AK048520.1 GI:26339371  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Izawa, M., Ohara, E., Tanaka, T., Matsura, S., Kawai, J., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipette sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE  
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehlin, P., Lewis, S., Mateu, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazza, R., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & II team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

6 (bases 1 to 3052)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatake, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/

FEATURES  
source

1..3052  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:C130068P22"  
/db\_xref="taxon:10090"  
/clone="C130068P22"  
/tissue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="16 days embryo"  
554..2381

CDS  
/note="unnamed protein product; fucosyltransferase 8 (MGD|MG1:1858901, GB|NM\_016893, evidence: BLASTN, 99%, match=2175)  
putative"

/codon\_start=1  
/protein\_id="BAC33357.1"  
/db\_xref="GI:26339372"  
/translation="MRAWTGSRWIMILFAMWGLTLLFYIGHLVLDNDHPHSHRELS  
KILAKRLKQONEDLRMAESLRIPGPIDQGTATGRVLEQLVKAKEQIENYKK  
QARNLQKDHILRRRTENGAKELWFFLOSELKLEGLNEHRADEILDLGHHE  
RSTMDLYLSQTDGAGDREKAKDLTEIVORITVYLPNDKDCSKARKLVNKGK  
GYGQQLHHVYCFMIAJQDRTORTILLESQNRWYATGGWETVPRPSETCTDRSGSTGH  
WSGENDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLURVSGDPAWVWSQFVKY  
LIRPQWLKEIEEATKGLGFKGVIGVHVRRTDKVGEAFHPIEETVMVHVEHLFQ  
LARRMQVQKRVILATDDPTLLKEAKTKYSNIEFISDNISSWSAGLHNRITYENLGV  
ILDHFLSQADFLVCTPSSOVCVAVYIMOTLHPDASANPHSLDDIYVFGQNAHQI  
AVTPHKPTEETPEPMDIIGVAGNHWGYSKGINRKLKGTGLYPSYKVKREKJETVK  
IPIYPEAKH"  
polyA\_signal 3036..3041  
/note="putative"  
polyA\_site 3052  
/note="putative"  
BASE COUNT 895 a 685 c 684 g 788 t  
ORIGIN

Query Match 90.1%; Score 1809.4; DB 11; Length 3052;

Best Local Similarity 94.1%; Pred. No. 0;

Matches 1891; Conservative 0; Mismatches 116; Indels 2; Gaps 1;

QY 1 AAGAAACCTTTATTTCTGTGTGCTAACTAGAACACAGAGTACAAATGTTTCCATTTCTT 60  
DB 553 AAAAGAAACCTTTATTTCTGTGTGCTAACTAGAACACAGAGTACAAATGTTTCCAGTTCTT 612  
QY 61 TGAGCTCCGAGACACAGACA -- AGGAGTGTGAACTCTGAAATCGGGCATCGACTGCTT 118  
DB 613 TGAGCTCAGGAGATGAGGACAGAGTTGAACTCTGAAATCGGGCATCGACTGCTT 672  
QY 119 CCTGCGGTGGATTATGCTCATCTTTTCTGCTGGGACCTTATTTGTTTATATAGGTG 178  
DB 673 CCTGCGGTGGATTATGCTCATCTTTTCTGCTGGGACCTTCTGTTTATATAGGTG 732  
QY 179 GTCAATTTGGTTCGAGATAATGACACACCTGACCAATCTTAGCAGAGAACTCTCCAAGATTC 238  
DB 733 GTCAATTTGGTTCGAGATAATGACACACCTGACCACTCCAGCAGAGAACTCTCCAAGATTC 792  
QY 239 TTGCAAGCTGGAGCGCTTAAACACAAATGAAGACTTCGAGGAGATGCTGAGTCTC 298  
DB 793 TTGCAAGCTTGAACGCTTAAACACAAATGAAGACTTCGAGGAGATGCTGAGTCTC 852  
QY 299 TCCGAATACCAAGAGGCGCTTATTTGATCAGGGGACAGCTTACAGGAGAGTCCGCTGTTTATG 358  
DB 853 TCCGAATACCAAGAGGCGCTTATTTGATCAGGGGACAGCTTACAGGAGAGTCCGCTGTTTATG 912  
QY 359 AAGAACAGCTTTAAGCCCAAAGACAGATTGAAATTTACAGAAACAAGCTAGGAATG 418  
DB 913 AAGAACAGCTTTAAGCCCAAAGACAGATTGAAATTTACAGAAACAAGCTAGGAATG 972  
QY 419 ATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGTGTGAAATTTGAGACTTAAAGAGCTCT 478  
DB 973 GTCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGTGTGAAATTTGAGACTTAAAGAGCTCT 1032  
QY 479 GGTTTTTTCTACAAGAGTGAATTTGAAGAAATTTAAGAAATTTAAGAGGAAACGAACTCCAAA 538  
DB 1033 GGTTTTTTCTACAAGAGGAGTGAATTTGAAGAAATTTAAGAGGAAATTTAAGAGGAAATTTAAGAGGAGTGAATTTCTATCATGACAGATC 1092  
QY 539 GACATGACAGATGAATTTCTTTTGGATTAGGACATCATGAAGTCTCTATCATGACAGATC 598  
DB 1093 GACATGACAGATGAATTTCTTTTGGATTAGGACATCATGAAGTCTCTATCATGACAGATC 1152  
QY 599 TATCTACCTCAGTCAAAACAGATGAGCAGGTGAGTGGCGGGAAGAAAGCAAGCAAGATC 658  
DB 1153 TATCTACCTCAGTCAAAACAGATGAGCAGGTGAGTGGCGGGAAGAAAGCAAGCAAGATC 1212  
QY 659 TGACAGAGCTGGTCCAGCGGGAATAACATATCTGCAGAAATCCCAGAGATCTCCAGCAAG 718  
DB 1213 TGACAGAGCTGGTCCAGCGGGAATAACATATCTCCAGAAATCCTTAAGAGCTGCAGCAAG 1272





|||||  
121 ATTCAAGTGGTCCAGCTCCCACTTTAGACAGCCTCCATCTCGGCCTCTTACTATCA 180  
QY  
1015 TTGGCTGTACCAAGACCTTCAGATCGACTCTCGAGAGTCCATGGTGTAGCTTCAGTG 1074  
Db  
181 CTGGCTGTTCAGAAAGACCTTCAGACCGACTCTTAAGAGTCCATGGTGGACCTGCA 240  
QY  
1075 TGGTGGGTATCCAGCTTTGTCAAAATCTTTGATCGGTCCACACCTTGGCTGGAAAGG 1134  
Db  
241 TGGTGGGTGTCCAGCTTTGTCAAAATCTTTGATCGGTCCACACCTTGGCTGGAAAGG 300  
QY  
1135 ATAGAAGAAACACCAAGAGCTTGGCTTCAAAATCTTCAAGTATTTGGAGTCCATGT 1194  
Db  
301 ATAGAAGAAACACCAAGAGCTTGGCTTCAAAATCTTCAAGTATTTGGAGTCCATGT 360  
QY  
1195 CGCACTGCAAAAGTGGGAACAGACAGCCTTCCATCCATTCAGAGTGAATACATGTG 1254  
Db  
361 CGCAGACAGAAAGTGGGAACAGACAGCCTTCCATCCATTCAGAGTGAATACATGTG 420  
QY  
1255 GTTGAAGAACATTTTTCAGCTTCTCGAAGCGAGAAATGAAAGTGGATTAATAAGAG 1314  
Db  
421 GTTGAAGAACATTTTTCAGCTTCTCGAAGCGAGAAATGAAAGTGGATTAATAAGAG 480  
QY  
1315 CTGGCCACTGATGACCTTCTTTGTTAAAGAGGAGGAAAGACAAAGTACTCCATTTAT 1374  
Db  
481 CTGGCTACTGATGATCTTCTTTGTTAAAGAGGAGGAAAGACAAAGTACTCCATTTAT 540  
QY  
1375 TTTATTAGTGATACTTCTTTTGGTTCAGCTGAGTACACCAACCGATACACAGAAAT 1434  
Db  
541 TTTATTAGTGATACTTCTTTTGGTTCAGCTGAGTACACCAACCGATACACAGAAAT 600  
QY  
1435 TCATCTTCGGGGCGTGATCTCGATATACATTTTCTCTCCAGGCTGACTTCTCTGTGT 1494  
Db  
601 TCATCTTCGGGGCGTGATCTCGATATACATTTTCTCTCACAGCTGACTTCTCTGTGT 660  
QY  
1495 ACTTTTTCATCCAGCTCTGTAGGTTGCTTATGAAATCATGCAAACTGATCCTGAT 1554  
Db  
661 ACTTTTTCATCCAGCTCTGTAGGTTGCTTATGAAATCATGCAAACTGATCCTGAT 720  
QY  
1555 GCCTCTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCAC 1614  
Db  
721 GCCTCTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGAGGCAAAATGCCAC 780  
QY  
1615 AACAGATTGCAATTTATCTCCCAACCTCGAATTAAGAGGAAATCCCATGGAACCT 1674  
Db  
781 AATCAGATTGCTGTTTATCTCCCAACCTCGAATTAAGAGGAAATCCCATGGAACCT 840  
QY  
1675 GGAGATATCATTTGCTGCTGGAACCT-ATTGGAATGTTTACTTAAAGGTGTCACAG 1733  
Db  
841 GGAGATATCATTTGCTGCTGGAACCTAATGGGATGTTTATTTCTAAAGGTATCAACAG 900  
QY  
1734 AAACTAGGAAAAACAGGCTCTACCTTCC 1764  
Db  
901 AAACTAGGAAAAACAGGCTTATATCCCTTC 931

RESULT 4  
BM928521  
LOCUS  
DEFINITION  
AGENCOURT\_6715254 NIH\_MGC\_100 Homo sapiens cdna clone IMAGE:5798262  
5', mRNA sequence.  
ACCESSION  
BM928521  
VERSION  
BM928521.1 GI:19378900  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1066)  
AUTHORS  
NIH-MGC <http://mgc.ncl.nih.gov/>  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.

Email: csapbe-re@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cdna Library Preparation: Rubin Laboratory  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LiCM2025 row: p column: 07  
High quality sequence stop: 666.

FEATURES  
source  
Location/Qualifiers  
1..1066  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:5798262"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_100"  
/notes="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cdna made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCACGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 315 a 235 c 248 g 264 t 4 others

Query Match 40.3%; Score 810; DB 12; Length 1066;  
Best Local Similarity 90.7%; Pred. No. 5.7e-182;  
Matches 894; Conservative 0; Mismatches 89; Indels 3; Gaps 3;  
QY 465 AGCTAAAGAGCTCTGGTTTCTTACAAAGTGAATTCGAAGAAATTAAGAATTTAGAGG 524  
Db 1 AGCTAAAGAGCTCTGGTTTCTTACAGAGTGAATTAAGAATTTAGAGAACTTTAGAGG 60  
QY 525 AAACGAACTCCAAAGACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAAGGTC 584  
Db 61 AATGAACTCCAAAGACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAAGGTC 120  
QY 585 TATCATGACAGATCTATACCTCAGTCAAAAGATGGAGCAGGTGAGTGGCGGAAAA 644  
Db 121 TATAATGCGATCTATACCTCAGTCAAGACAGATGGAGCAGGTGATTGGCGGAAAA 180  
QY 645 AGAGCCAAAGATCTGACAGAGCTGGTCCAGCGAGAAATACATATCTCGAGAATCCCAA 704  
Db 181 AGAGCCAAAGATCTGACAGAGCTGGTTCAGCGAGAAATACATATCTTCAGAATCCCAA 240  
QY 705 GGACTGACGAAAGCCAGAGAGCTGGTATGTAATATCAAAAGGCTGTGGCTATGGATG 764  
Db 241 GGACTGACGAAAGCCAAAGAGCTGGTATGTAATATCAAAAGGCTGTGGCTATGGCTG 300  
QY 765 TCAACTCCATCATGTGGTTTACTCTCTCATGATTTATGGCACCAGCGAACTCAT 824  
Db 301 TCAGTCCATCATGTGGTCTACTCTCTCATGATTTGTCATATGGCACCAGCGAACTCAT 360  
QY 825 CTTGGAATCTCAGAAATTTGGCGCTATGCTACTGGAGGATGGAGAGCTGTGTTTAGACCTGT 884  
Db 361 CTTGGAATCTCAGAAATTTGGCGCTATGCTACTGGTGGATGGGAGACTGTATTTAGCCCTGT 420  
QY 885 AAGTGAGACATGCACAGACAGCTTGGCCTCTCCACTGGACACTGGTTCAGGTGAAGTGA 944  
Db 421 AAGTGAGACATGCACAGACAGATCTGGCATCTCCACTGGACACTGGTTCAGGTGAAGTGA 480  
QY 945 GGACAAATAATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTCCATCTCTGCTCTCC 1004  
Db 481 GGACAAATAATTTCAAGTGTGAGCTTCCCATTTGTAGACAGCTTCTTATCCCGCTCTCC 540  
QY 1005 TTACTTACCTTGGCTGTGTACCAGAGACCTTTCAGATCGACTCTCTGAGGTCCATGGTGA 1064

Db 541 ATATTACCTTGGGTGTACACAGAGACCTCGCAGATCGACTGTGTACAGTGCATGGTGA 600  
QY 1065 TCCTCAGTGTGGTGTATCCAGTGTGTCAATACCTTGTATCCGTTCACAACTTTGGCT 1124  
Db 601 CCTCAGTGTGGTGTCTCAGTGTGTCAATACCTTGTATCCGTTCACAACTTTGGCT 660  
QY 1125 GGAAGGGAATPAGAAGAACCAACCAAGAGCTTGGCTTCAAACTCAGTATTGGAGT 1184  
Db 661 AGAAANAGAAATPAGAAGAGCCACCAAGAGCTTGGCTTCAAACTCAGTATTGGAGT 720  
QY 1185 CCATGTACAGCCTGTCAAAAGTGGGACAGAACAGCTTCCATCCATGGAGATA 1244  
Db 721 CCATGTACAGCCTGTCAAAAGTGGGACAGAACAGCTTCCATCCATGGAGATA 780  
QY 1245 CATGTACAGCTTGAAGAACATTTTTCAGCTTCTCGAACGAGAAATGAAGTGAATAAAA 1304  
Db 781 CATGTGATTTTGAAGAACATTTTTCAGCTTCTCGAACGAGAAATGCNAGTGGACANAA 840  
QY 1305 AAGAGTGTATCTGGCCTGTATGACCTTTCTTTTAAAGGAGGCAAAAGTACTC 1364  
Db 841 AAGAGTGTATTTGGCCACAGATGACCTTTCTTTTAAAGGAGGCAAAAGTACTC 900  
QY 1365 CAATTATGAATTT-ATTAGTGATTAATCTATTTT-CTTGGTACCTGGAC-TACACAACGG 1421  
Db 901 CAATTATGAATTTTAAATGAATCTATTTTCCCTGTCTGTCAGCGGGACTTGGCCATCGN 960  
QY 1422 ATACACAGAAATTCACCTCGGGGG 1447  
Db 961 ATACCCCGAAATTCCTTCCGGGG 986

RESULT 5  
LOCUS BQ653329  
DEFINITION AGENCOURT\_8298428 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6270458  
5', mRNA sequence.  
ACCESSION BQ653329  
VERSION BQ653329.1 GI:21777501  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>  
1 (bases 1 to 940)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM2445 row: k column: 03  
High quality sequence stop: 700.  
Location/Qualifiers  
1..940  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6270458"  
/tissue\_type="Hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 100"  
/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)

FEATURES  
source

using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 286 a 197 c 222 g 235 t

ORIGIN

Query Match 38.2%; Score 768; DB 13; Length 940;  
Best Local Similarity 92.0%; Pred. No. 5.6e-172;  
Matches 865; Conservative 0; Mismatches 70; Indels 5; Gaps 5;

QY 465 AGCTAAAGAGCTCTGGTTTTTCTACAAAGTGAATTAAGAAATTAAGAAATTAAGAGG 524  
Db 1 AGCTAAAGAGCTCTGGTTTTTCTACAGAGTGAATTAAGAAATTAAGAAATTAAGAGG 60  
QY 525 AAACGAATCTCAAGACATGACAGATGAATTTCTTTTGGATTAGACATCATGAAGGTC 584  
Db 61 AAATGAATCTCAAGACATGACAGATGAATTTCTTTTGGATTAGACATCATGAAGGTC 120  
QY 585 TATCATGACAGATCTATACCTCAGTCAAAAGATGAGCAGGTGAGTGGCGGAAAA 644  
Db 121 TATAATGAGGATCTATACCTCAGTCAAGATGAGCAGGTGAGTGGCGGAAAA 180  
QY 645 AGAAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTGCAAGATCCAA 704  
Db 181 AGAGGCCAAAGATCTGACAGAACTGGTTTCAGCGGAGAAATACATATCTTCAAGATCCAA 240  
QY 705 GGACTGCGAGCAAGCCAGAAAGCTGGTATGTAATATCAACAAAGCTGTGGCTATGGATG 764  
Db 241 GGACTGCGAGCAAGCCAGAAAGCTGGTATGTAATATCAACAAAGCTGTGGCTATGGATG 300  
QY 765 TCAACTCCATCATGTGTGTACTGCTTCATGATTGCTTATGGCACCCAGCGAAACATCAT 824  
Db 301 TCAGCTCCATCATGTGTGTACTGCTTCATGATTGCAATATGGCACCCAGCGAAACATCAT 360  
QY 825 CTTGGAATCTCAGAAATTGGCGCTATGCTACTGAGGAGTGGAGACTGTGTTTAGACCTGT 884  
Db 361 CTTGGAATCTCAGAAATTGGCGCTATGCTACTGAGGAGTGGAGACTGTGTTTAGACCTGT 420  
QY 885 AAGTGAGACATGCACAGACAGTCTGGCTCTCCACTGGACACTGGTCAGGTGAAGTGA 944  
Db 421 AAGTGAGACATGCACAGACAGTCTGGCTCTCCACTGGACACTGGTCAGGTGAAGTGA 480  
QY 945 GGACAAAATATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTTCCATCTCGTCTCC 1004  
Db 481 GGACAAAATATTTCAAGTGTGAGCTCCCATTTGTAGACAGCTTCCATCTCGTCTCC 540  
QY 1005 TTACTTACCTTGGCTGTACAGAGACTTGGCAGATCGACTCCTGAGAGTCCATGGTGA 1064  
Db 541 ATATTTACCTTGGCTGTACAGAGACTTGGCAGATCGACTTGTACGAGTGCATGGTGA 600  
QY 1065 TCCTGCACTGTGGTGTATCCAGTTTGTCAATACTTGTATCCGTCCACAACTTTGGCT 1124  
Db 601 CCCTGCACTGTGGTGTATCCAGTTTGTCAATACTTGTATCCGTCCACAACTTTGGCT 660  
QY 1125 GGAAGGGAATPAGAAGAACCAACCAAGAGCTTGGCTTCAAACTCAGTATTGGAGT 1184  
Db 661 AGAAANAGAAATPAGAAGAGCCACCAAGAGCTTGGCTTCAAACTCAGTATTGGAGT 719  
QY 1185 CCATGTACAGCCTGTGACAAAGTGGGACAGAACAGCTTCCATCCATGGAGATA 1244  
Db 720 CCATGTACAGCCTGTGACAAAGTGGGACAGAACAGCTTCCATCCATGGAGATA 779  
QY 1245 CATGGT-ACAGTTGAAGAACTTTTTCAGCTTCTCGAACGAGAAATGAAGTGAATA 1303  
Db 780 CATGGTGAAGTGTGAAGAACTTTTTCAGCTTCTCGAACGAGAAATGAAGTGAATA 839  
QY 1304 AAAGAGTGA-TCTGGCCACTGTAGCCCTTCTTTGT-TAAAGGAGGCAAAAGCAAGTA 1361  
Db 840 AAAGAGTGA-TTTAGTGAATGACCTTCTTTTATTAAGGAGGCAAAACCAAGTA 899  
QY 1362 CTCCTAATATGAA-TTTATAGTGAATGACCTTCTTTTCTTTTCTTTTCTTTTCTTTG 1400  
Db 900 CCCCCATTTATGAATTTTATTTATGATTAATCTTCTTTTCTTTTCTTTTCTTTTCTG 939

```
RESULT 6
BX452321
LOCUS
DEFINITION
BX452321 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CS0D6005Y007 5-PRIME, mRNA sequence.
ACCESSION
BX452321
VERSION
BX452321.1 GI:31026371
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7499.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAG023ZH04_CS02144_1&cluster=7499.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG023ZH04_CS02144_1.
Location/Qualifiers
1. .900
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D6005Y007"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 250 a 201 c 200 g 245 t 4 others
ORIGIN
Query Match 37.7%; Score 757.6; DB 13; Length 900;
Best Local Similarity 91.6%; Pred. No. 1.7e-169;
Matches 810; Conservative 0; Mismatches 73; Indels 1; Gaps 1;
QY 742 AACAAAGGCTGGGTATGGATGTCAACTCCATCATGTGGTTTACTGCTTCATGATTCT 801
DB 15 AACAAAGGCTGGGTATGGCTGTGAGCTCCAT-ATGGTGCTACTGCTTCATGATGCA 73
QY 802 TATGGCACCACGGAACACTCATCTTGAATCTCAGAAATTCGAGCTGCTACTGAGGA 861
DB 74 TATGGCACCACGGAACACTCATCTTGAATCTCAGAAATTCGAGCTGCTACTGAGGA 133
QY 862 TGGGAGACTGTTTATAGCTCTTAAGTAGAGACATGCACAGAGCTGCGCTTCCTACT 921
DB 134 TGGGAGACTGTTTATAGCTCTTAAGTAGAGACATGCACAGAGCTGCGCTTCCTACT 193
QY 922 GGACACTGTCAGGTGAAGTGAAGACAAAATGTTCAAGTGGTGGCTCCCATCTGTA 981
DB 194 GGACACTGTCAGGTGAAGTGAAGACAAAATGTTCAAGTGGTGGCTCCCATCTGTA 253
QY 982 GACAGCTCCATCTCGTCTCTCTTACTTACCTTGGCTGTACCAAGAGACTTGCAGAT 1041
DB 254 GACAGCTCTTATCCCGCTCTCCATATTTACCTTGGCTGTACCAAGAGACTTGCAGAT 313
QY 1042 CGACTCCTGAGAGTCATCGTGTATCTCGAGTGGTGGTATCCAGTTGTCAATATAC 1101
```

```
DB 314 CGACTTGTACAGTGCATGGTGAACCCCTGCAGTGGTGGTGTCTCAGTTGTCAATATAC 373
QY 1102 TTGATCCGTCCAAACCTTGGCTGGAAAGGGAATAGAGAAACCAACCAAGAACTTGGC 1161
DB 374 TTGATCCGCCACAGCTTGGCTAGAAAAGAAATAGAGAAGCCACCAAGAACTTGGC 433
QY 1162 TTCAAAATCCAGTTATGGATCCATGTAGAGCCACTGACAAAGTGGGACAGAGCA 1221
DB 434 TTCAAAATCCAGTTATGGATCCATGTAGAGCCACTGACAAAGTGGGACAGAGCT 493
QY 1222 GCCTTCCATCCATTGGAGGAATACATGTACACCTTGAAGAAACATTTTTCAGCTTCTCGAA 1281
DB 494 GCCTTCCATCCATTGGAGGAATACATGTAGAGCCACTGACAAAGTGGGACAGAGCA 553
QY 1282 CGCAGAATGAAGTGGATATAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTA 1341
DB 554 CGCAGAATGAAGTGGATATAAAAGAGTGTATCTGGCCACTGATGACCCCTTCTTTGTA 613
QY 1342 AAGGAGCAAGAACAAAGTACTCCAAATATGAATTTATAGTATTAATCTTCTTCTGG 1401
DB 614 AAGGAGCAAGAACAAAGTACTCCAAATATGAATTTATAGTATTAATCTTCTTCTGG 673
QY 1402 TCAGCTGGACTACACAAACCGATACACAGAAATTCACCTTGGGGCGTGATCCTCGATATA 1461
DB 674 TCAGCTGGACTACACAAACCGATACACAGAAATTCACCTTGGGGCGTGATCCTCGATATA 733
QY 1462 CACTTCTCTCCAGGCTGACTTCTCTTGTGTGTAATTTTTCATCCAGGCTGTAGGGTT 1521
DB 734 CATNTTCTCTCAAGCAGACTTCTTCTAGTGTGTAATTTTTCATCCAGGCTGTAGGGTT 793
QY 1522 GCTTATGAATCATGCAACACATGCTATCCTGATGCCTCTGCAAACTTCCATCTTTAGAT 1581
DB 794 GCTTATGAATCATGCAACACATGCTATCCTGATGCCTCTGCAAACTTCCATCTTTAGAT 853
QY 1582 GACATCTACTATTTTGGAGGCCAAATGCCACACAGATTGC 1625
DB 854 GACATTTACTATTTTGGGGCCAGATNGCCCAATCAAAATTC 897
RESULT 7
BX423860
LOCUS
DEFINITION
BX423860 886 bp mRNA linear EST 23-MAY-2002
AGENCOURT 7894821 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6158200
5', mRNA sequence.
ACCESSION
BX423860
VERSION
BX423860.1 GI:21119175
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTb/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13505 row: e column: 17
High quality sequence stop: 641.
Location/Qualifiers
1. .886
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6158200"
/tissue_type="melanotic melanoma"
```

```
/lab host="DH10B (phage-resistant)"
/clone lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      273 a 185 c 212 g 215 t      1 others
ORIGIN

Query Match      37.4%; Score 751.2; DB 13; Length 886;
Best Local Similarity 91.9%; Pred. No. 5.6e-165;
Matches 804; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

QY 507 ATTAAGAAATAGAGGAAACGAACCTCAAGACATGCAGATGAAATCTTTTGGATTT 566
Db 1 ATTAAGAACTTAGAAGGAATGAACTCAAGACATGCAGATGAAATCTTTTGGATTT 60

QY 567 AGGACATCATGAAGGTCTATCATGACAGATCTATACCTCACTCAAGCATGGAGC 626
Db 61 AGGACATCATGAAGGTCTATAATGACGGATCTATACCTCACTCAAGCATGGAGC 120

QY 627 AGGTGAGTGGCGGAAAGAGCAAGATCTGACAGAGTGGTCCAGCGGAGATAAC 686
Db 121 AGGTGATTTGGCGGAAAGAGCGCAAGATCTGACAGAGTGGTTCAGCGGAGATAAC 180

QY 687 ATATCTGACAGATCCCAAGGACTGCAGCAAGCCAGAAAGCTGGTATGTATATCAACA 746
Db 181 ATATCTTCAGATCCCAAGGACTGCAGCAAGCCAGAAAGCTGGTATGTATATCAACA 240

QY 747 AGGTGTTGGCTATGATGTCAATCCATCATGTGTGTTTACTGCTTCATGATGCTTATGG 806
Db 241 AGGTGTTGGCTATGCTGTGCTGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

QY 807 CACCAGCGAAGACACTCATCTTGGATCTCAGAAATTTGGCGCTATGCTACTGGAGGTGGA 866
Db 301 CACCAGCGAAGACACTCATCTTGGATCTCAGAAATTTGGCGCTATGCTACTGGAGGTGGA 360

QY 867 GACTGTGTTAGACCTGTAAGTGACATGCACAGACAGGTCTGGCTCTCCACTGGACA 926
Db 361 GACTGTATTTAGGCTGTGATGACATGCACAGACAGTCTGGCTCTCCACTGGACA 420

QY 927 CTGGTCAGTGAAGTGAAGGACAAAGATTTCAAGTGTGCGAGTCCCAAGTGTAGACAG 986
Db 421 CTGGTCAGTGAAGTGAAGGACAAAGATTTCAAGTGTGCGAGTCCCAAGTGTAGACAG 480

QY 987 CTTCCATCTCTGCTCTCTTACTTACCTTGGCTGTACAGAGACCTTGCAGATCGACT 1046
Db 481 TCTTCATCCCGTCTCTCCATATTTTACCTTGGCTGTACAGAGACCTTGCAGATCGACT 540

QY 1047 CTTGAGAGTCCATGTTGATCTCTGAGTGTGGGTATGCCAGTGTGTCAAATCTTCAT 1106
Db 541 TGTACAGTGCATGTTGATGACCTTGCAGTGTGGGTGTCTCAGTTGTCAATATCTGAT 600

QY 1107 CGGTCCACACCTTGGCTGGAAAGGAAATAGAAGAAACCAAGAGAGCTTGGCTTCAA 1166
Db 601 CGGCCCCACAGCTTGGCTAGAAAAAGAAATAGAAGAGCCCAAGAGAGCTTGGCTTCAA 660

QY 1167 ACATCCAGTTATGAGTCCATGTGACAGCGACTCAGAAAGTGGGAAACAGAGAGCCTT 1226
Db 661 ACATCCAGTTATGAGTCCATGTGACAGCGCAAGAGTGGGAAACAGAGAGCCTT 720

QY 1227 CCATCCCATTTAGGAATACATGTTACATGTTGAGGATTTTTCAGCTTCTCGAACGCGAG 1286
Db 721 CCATCCCATTTAGGAATACATGTTGAGGATTTTTCAGCTTCTCGAACGCGAG 780

QY 1287 AATGAAGTGGTAAAAAGAGTGTATCTGCCCACTGATGACCC - TTTCTTTGTTAAAG 1344
Db 781 AATGAAGTGGTAAAAAGAGTGTATTTGCCCACTGATGATGACCCCTTTTAAAG 840

QY 1345 GAGGCAAGACAAAGTACTTCCAAATTTAT 1379
Db 841 GAGGCAAGACAAAGTACTTCCAAATTTAT 875
```

## RESULT 8

BX429921

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7499.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0BAA0122A08\_CS01114\_1&amp;cluster=7499.f.

Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0BAA0122A08\_CS01114\_1.

Location/Qualifiers

1. 923

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D0051007"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6; 1st strand CDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand CDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 285 a 184 c 231 g 218 t 5 others

ORIGIN

Query Match 36.9%; Score 741; DB 13; Length 923;

Best Local Similarity 92.5%; Pred. No. 1.5e-165;

Matches 821; Conservative 0; Mismatches 63; Indels 4; Gaps 4;

QY 423 GGGAAAGGATCATG-AAATCTTAAGGAGGAGGATTGAAATGGAGCTTAAGAGCTCTGGT 481

Db 31 GGGAAAGGATCATGAAATCTCTGAGGAGGAGGATTGAAATGGAGCTTAAGAGCTCTGGT 90

QY 482 TTTTCTTACAAAGTGAATTTGAAGAAATTAAGAAATTTAGAGAGAAACGAACCTCCAAGAC 541

Db 91 TTTTCTTACAGTGAATTTGAAGAAATTAAGAAATTTAGAGAGAAATGAATCTCCAAGAC 150

QY 542 ATGAGATGAATTTCTTTTGGATTTAGACATCATGATGAAGGTCTATCATGACATCTAT 601

Db 151 ATGAGATGAATTTCTTTTGGATTTAGACATCATGATGAAGGTCTATATGATCGGATCTAT 210

QY 602 ACTACCTCAGTCAACAGATGGAGCGAGTGGTGGCGGAAAGAGCCAAAGATCTGA 661

Db 211 ACTACCTCAGTCAACAGATGGAGCGAGTGGTGGCGGAAAGAGCCAAAGATCTGA 270

QY 662 CAGAGCTGGTCCAGCGAGGATTAACATATCTGACAGATTTCCCAAGAGCTGCGAGCAAGCCA 721

Db 271 CAGAGCTGGTCCAGCGAGGATTAACATATCTTCAAGATTTCCCAAGAGCTGCGAGCAAGCCA 330

QY 722 GAAAGCTGGTATGTAATATCAACAAAGCTGTGGCTATGGAATGTCAACTCCATCATGTGG 781

Db 331 AAAAGCTGGTGTGTAATATCAACAAAGCTGTGGCTATGGAATGTCAGCTCCATCATGTGG 390

```
QY 782 TTTACTGCTTCTATGATTGCTTATGACCCAGCAACATCTCATCTTGAATCTCAGAATT 841
Db 391 TCTACTGCTTCTATGATTGCTATGTCATATGTCACCCAGCAACATCTCATCTTGAATCTCAGAATT 450
QY 842 GCGCTATGCTACTGAGATGGGAGACTGTTGTTAGACCTGTAGTACGATCAGACATCAG 901
Db 451 GCGCTATGCTACTGCTGAGTGGGAGACTGTATTTAGGCTGTAAAGTGGAGATCAGACAG 510
QY 902 ACAGTCTGGCTCTCCACTGACACTGCTCAGTGAAGTGAAGGACAAAAATGTTCAAG 961
Db 511 ACAGATCTGGCATCTCCACTGACACTGCTCAGTGAAGTGAAGGACAAAAATGTTCAAG 570
QY 962 TGGTGGAGTCCCTTGTAGACAGCTCCATCTCGTCTCTTACTTACCTTTGGCTG 1021
Db 571 TGGTGGAGTCCCTTGTAGACAGCTTTCATCCCGTCTCTCATATTTACCTTTGGCTG 630
QY 1022 TACCAAGAACCTTGCAGATCGACTTCCTGAGATCCATGGTGATCCTGCAGTGGTGG 1081
Db 631 TACCAGAAAGCTCGCAGATCGACTTGTACGAGTGCATGGTGACCTGCAGTGGTGG 690
QY 1082 TATCCAGTTTGTCAAACTTGTATCGTCCCAACCTTGGCTGAAAGGAAATAGAA 1141
Db 691 TGTCTCAGTTTGTCAAACTTGTATCGTCCCAACCTTGGCTGAAAGGAAATAGAA 749
QY 1142 AAACACCAAGAGCTTGGCTTCAAACTTGTATCGTCCCAACCTTGGCTGAAAGGAAATAGAA 1201
Db 750 AAGCCACCAAGAGCTTGGCTTCAAACTTGTATCGTCCCAACCTTGGCTGAAAGGAAATAGAA 808
QY 1202 ACAAGTGGGAAACAGAGCAGCTTCCATCCATTCATTTAGGATCCATGTCAGACGCACTG 1260
Db 809 ACAAGTGGGAAACAGAGCAGCTTCCATCCATTCATTTAGGATCCATGTCAGACGCACTG 868
QY 1261 GAACATTTTCACTTCTGACGCGAGATGGAAGTGGATTAAGAA 1308
Db 869 GAACATTTTCACTTCTGACGCGAAATGCAAGTGGAGGACAAAGGA 916
```

```
RESULT 9
LOCUS BX450010 1027 bp mRNA linear EST 22-MAY-2003
DEFINITION BX450010 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CS0DG005Y007 5-PRIME, mRNA sequence.
```

```
ACCESSION BX450010
```

```
VERSION BX450010.1 GI:31032111
```

```
KEYWORDS EST.
```

```
SOURCE Homo sapiens (human)
```

```
ORGANISM
```

```
REFERENCE
```

```
AUTHORS
```

```
TITLE
```

```
JOURNAL
```

```
COMMENT
```

```
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7499.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
```

```
cgi-bin/cluster.cgi?seq=CS0A5009ZB08QP1&cluster=7499.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0A5009ZB08QP1.
```

```
FEATURES
```

```
source
```

```
1..1027
```

```
/organism="Homo sapiens"
```

```
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
```

```
/clone="CS0DG005Y007"
```

```
/tissue_type="B CELLS (RAMOS CELL LINE)"
```

```
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 320 a 188 c 258 g 252 t 9 others
ORIGIN
```

```
Query Match 36 9%; Score 741; DB 13; Length 1027;
```

```
Best Local Similarity 90.1%; Pred. No. 1.6e-165; Indels 2; Gaps 1;
```

```
Matches 801; Conservative 3; Mismatches 83;
```

```
QY 1 AACAGAAACCTTATTTCTGTGTGGCTAACTAGAACACAGAGTACAATGTTTCCAAATCTT 60
```

```
Db 139 AACAGAAAGTCTATTACCTGTGCACCTAACTAGAAACACAGAGTTACAATGTTTCCAAATCTT 198
```

```
QY 61 TGAGCTCCGAGAAGACA--GAAGGAGTTGAACTCTGAAATCGGAAATCGGCGATGGACTGTT 118
```

```
Db 199 TGAGCTCCAGGACTCCAGGGAAGTGAGTTGAAATCTGAAATCGGCGATGGACTGTT 258
```

```
QY 119 CCTGCGCTTGAATATGCTCATCTTTTGCCTCGGGGACCTTATTTGTTTATATAGTG 178
```

```
Db 259 CCTGCGCTTGAATATGCTCATCTTTTGCCTCGGGGACCTTCTGCTGTTTATATAGTG 318
```

```
QY 179 GTCAATTTGTTTCGAGATAATGACCACTGACCAATTTCTAGCAGAGAACTCTCCAAGATTC 238
```

```
Db 319 GTCAATTTGTTTCGAGATAATGACCACTGACCAATTTCTAGCAGAGAACTCTCCAAGATTC 378
```

```
QY 239 TTGCAAGCTGGAGCGCTTAAACACAAATGAAGACTTGAGGAGAAATGGAGTCTGAGTCTC 298
```

```
Db 379 TGGCAAGCTTGAACGCTTAAACACAGAGAATGAAGACTTTGAGGCGAATGGCCGAATCTC 438
```

```
QY 299 TCCGAATACCAAGAGCGCTTATTCATCAGGGGACAGCTACAGGAGAGTCCGCTGTTTATG 358
```

```
Db 439 TCCGATACCAAGAGCGCTTATTCATCAGGGGACAGCTATAGGAGAGTACGCGTGTATG 498
```

```
QY 359 AAGAACAGCTTGTTTAAGGCCAAAGAACAGATTTGAAATTTACAAGAAACAACTAGGAATG 418
```

```
Db 499 AAGAGAGCTTGTTTAAGGCCAAAGAACAGATTTGAAATTTACAAGAAACAGACCAGAAATG 558
```

```
QY 419 ATCTGGGAAAGGATCATGAAATCTTAAAGGAGGAGATTTGAAATGGAGCTTAAAGAGCTCT 478
```

```
Db 559 GTCTGGGGAAGGATCATGAAATCTTGAAGGAGGAGATTTGAAATGGAGCTTAAAGAGCTCT 618
```

```
QY 479 GGTTTTCTTACAAAGTGAAATTTGAAGAAATTTAAAGAAATTTAGAAAGGAAACCAACTCCAAA 538
```

```
Db 619 GGTTTTCTTACAGAGTGAAATTTGAAGAAATTTAAAGAACTTTAGAAAGAAATGAACTCCAAA 678
```

```
QY 539 GACATGCAGATGAAATTTCTTTTGGATTTAGGACATCATGAAAGGCTCTATATGACGATC 598
```

```
Db 679 GACATGCAGATGAAATTTCTTTTGGATTTAGGACATCATGAAAGGCTCTATATGACGATC 738
```

```
QY 599 TATACCTCTCAGTCAAAACAGATGGAGCTGAGTGGCGGGAAGAAAGAACCCAAAGATC 658
```

```
Db 739 TATACCTCTCAGTCAAAACAGATGGAGCTGAGTGGCGGGAAGAAAGAACCCAAAGATC 798
```

```
QY 659 TGACAGAGCTGCTCCAGCGGAGATACATATCTGCAGAAATCCCAAGAGCTGCAGCAAG 718
```

```
Db 799 TGACAGAACTGGTTCAGCGGAGAAATACATATCTTCAGAAATCCCAAGAGCTGCAGCAAG 858
```

```
QY 719 CCAGAAAGCTGGTATGTAATATCAAAAGAGCTGTGGCTATGGATGTCACTCCATCATG 778
```

```
Db 859 CCAGAAAGCTGGTATGTAATATCAAAAGAGCTGTGGCTATGGCTGTGCTCAGCTCCATCATG 918
```

```
QY 779 TGGTTACTGCTTCATGATGCTTATGGCAGCCAGCAACTCATCTTGAATCTCAGA 838
```

```
Db 919 TGGTCTACTGCTTCATGATGCTTATGGCAGCCAGCAACTCATCTTGAATCTCAGA 978
```

```
QY 839 ATTGGCGCTATGCTACTGGGAGGATGGGAGACTGTTTATAGACCTGTAAG 887
```

```
Db 979 ATTGGCGCTATGCTACTGGGAGGATGGGAGACTGTTTATAGCTGTGTRAG 1027
```





Normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5 and 15.5 (size selected for the 0.5-1 kb fragments) cloned directionally, priming method: Oligo-dT. cDNA enrichment: >1k bp, Average insert size 1.6k bp. Normalization (Cot value): 7.5 kb. Priming sequence: 5'-GAC/TAG/TCT/ATG/ATCGGACGGCGGCC/T3'. Tissue contributed by David Rowe. Library constructed by Resgen, Invitrogen Corp.

BASE COUNT	ORIGIN				Query Match Best Local Similarity Matches	36.4%; 93.4%; 784;	Score 730.2; DB 14; Pred. No. 5.6e-163; Conservative 0;	Mismatches 53;	Indels 2;	Gaps 2;	Length 898;
	272 a	203 c	193 g	230 t							
Qy	998	GTCTCTCTTACTTACCCCTTGGCTGTACACAGAACCTTGCAGATCGACTCTCTGAGAGTCC	1057								
Db	1	GGCTCTCTTACTTACCACTGGCTGTTCACAGAACCTTGCAGACCGACTCTCTAAAAGTCC	60								
Qy	1058	ATGCTGATCCTGCAGTGTGTGGTATCCACAGTTTGTCAAAATACTTGTATCCGTCACAAC	1117								
Db	61	ATGCTGACCTTCAGTGTGTGGTGTCCAGTTTGTCAAAATACTTGTATTCGTCCACAAC	120								
Qy	1118	CTTGGCTGGAAAGGGAATAGAAAGAACACCAAGAGCTTGGCTTCAAAACATCCAGTTA	1177								
Db	121	CTTGGCTGGAAAGGGAATAGAAAGACCCAAAGAGCTTGGCTTCAAAACATCCAGTTA	180								
Qy	1178	TTGAGTCCATGTTCAGACGCATGTACAAAGTGGGAACAGAGCGCTTCATCCCATTTG	1237								
Db	181	TTGAGTCCATGTTCAGACGCACACAAAGTGGGAACAGAGCGCTTCCACCCCATCG	240								
Qy	1238	AGGAATACATGGTACACGTTTGAAGAACATTTTCAGCTTTCGAAAGCGAGAAATCAAAAGTGG	1297								
Db	241	AGAGTACATGGTACACGTTTGAAGAACATTTTCAGCTTTCGCAAGCGAGAAATCGAAGTGG	300								
Qy	1298	ATAAAAAAGAGTGTATCTCGCCCACTGATCAACCCCTCTTTTGTAAAGGAGGCAAGACAA	1357								
Db	301	ATAAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTGTAAAGGAGGCAAGACAA	360								
Qy	1358	AGTACTCCAATATGAATTTTATAGTGAATCACTCTATTTCTCGTCAGCTGGACTACACA	1417								
Db	361	AGTACTCCAATATGAATTTTATAGTGAATCACTCTATTTCTTGGTCAGCTGGACTACACA	420								
Qy	1418	ACCGATACACAGAAAAATTCATCTCGGGGCGTGATCCTGGATATACACTTTCTCTCCACGG	1477								
Db	421	ATCGGTACACAGAAAAATTCATCTCGGGGTGTGATCCTGGATATACACTTTCTCTCACAGG	480								
Qy	1478	CTGACTTCCTTGTGTACTTTTTTCATCCAGGTCGTGTAGGGTTGCTATGAAATCATGC	1537								
Db	481	CTGACTTCTTGTGTACTTTTTTCATCCAGGTCGTGTAGGGTTGCTATGAAATCATGC	540								
Qy	1538	AAACACTGCATCCTGATGCCTCTGCAAACTTCCATTCCTTTAGATGACATCTACTATTTTCG	1597								
Db	541	AAACCTTGCATCCTGATGCCTCTGCAAACTTCCATTCCTTTGGATGACATCTACTATTTTCG	600								
Qy	1598	GAGGCCAAAATGCCCAACACAGATTGCAGTTTATCTCCAAACCTCGAACTAAAAGAGG	1657								
Db	601	GAGGCCAAAATGCCCAACATCAGATTGCTGTGTTATCTCAAAAACCTCGAACTGAAAGAGG	660								
Qy	1658	AAATCCCATGGAACCTGGAGATATCATTTGGTGTGGCTGGAAACCATTTGGAATGGTTACT	1717								
Db	661	AAATTCCAATGGAACTTGGAGATATCATTTGGTGTGGCTGGGAACATTTGGGATGGTTATT	720								
Qy	1718	CTAAAGTGTCAACAGAAAAACTAGAAAAAACAGGCGCTGTATCCCTTCCTACAAAAGTCCGAG	1777								
Db	721	CTAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAG	780								
Qy	1778	AGAAGATAGAAAAACAG-TCAAATACCCCTTACATATCCCTG- AAGCTGAAAAATAGAGATGGA	1834								
Db	781	AGAAGATAGAAAAACGTTCAAGTATCCACATATCTGAAAGCTCAAAAAATAGAGATTGA	839								

RESULT 12	FEATURES	BASE COUNT
AUI24128	sour	ORIGIN
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		

Query Match	34.3%;	Score 689.2;	DB 9;	Length 834;
Best Local Similarity	90.2%;	Pred. No. 3.3e+153;		
Matches 747;	Conservative 0;	Mismatches 80;	Indels 1;	Gaps 1;
Qy 1004	CTTACTTACCCCTTGGCTGTACCAAGAAGACCTTGCAGATCGACTCTCGAGAGTCCATGGTG	1063		
Db 1	CATATTTACCCCTTGGCTGTACCAAGAAGACCTCGAGATCGACTTGTACGAGTGCATGGTG	60		
Qy 1064	ATCCTCGAGTGTGGTGGGTATCCACGTTTGTCAAACTTTGATCCGTCACCAACCTTGGC	1123		
Db 61	ACCTCGAGTGTGGNGGGTGTCTCAGTTTGTCAAACTTTGATCCGCCACACCCTTGGC	120		
Qy 1124	TGAAAGGGGAAATAGAAAGAAACCAACAAGAGCTTGGCTTCAAACATCCAGTTATTGGAG	1183		
Db 121	TAGAAAGGAAATAGAAAGAACCCACCAAGAGCTTGGCTTCAAACATCCAGTTATTGGAG	180		
Qy 1184	TCCATGTACAGCGCATGACAAAGTGGGAACAGAACAGCGCTTCCATCCCATTTGAGGAAT	1243		
Db 181	TCCATGTACAGCGCAGACAAAGTGGGAACAGAACGTGCCTTCCATCCCATTTGAAGAGT	240		
Qy 1244	ACATGTTACAGTTTGAAGAACATTTTTCAGCTTCTCGAACGCGAAATGAAAGTGGATAAAA	1303		
Db 241	ACATGTTGCATGTTTGAAGAACATTTTTCAGCTTCTTCGACGAGAAATGCAAGTGCACAAA	300		
Qy 1304	AAAGAGTGTATCTGGCCACTGTAGACCCCTTCTTTTGTAAAGGAGGCAAAACAAAGTACT	1363		
Db 301	AAAGAGTGTATTTTGGCCACAGATGACCTTCTTTTAAAGGAGGCAAAACAAAGTACT	360		



1364	QY	CCAAATTATGAATTTATTTAGTGATACATCTCTATTTCTTGGTCAGCTGGAGCTACACACCGAT	1423
361	Db	CCAAATTATGAATTTATTTAGTGATAAATCTATTTCTTGGTCAGCTGGAGCTGCACAATCGAT	420
1424	QY	ACACAGAAAAATTCACATTCGGGGGCGTGATCTGGGATATACACTTTTCTCTCCACAGGCTGACT	1483
421	Db	ACACAGAAAATTCATCTCGTGGAGTGATCTGGATATACATTTTCTCTCAGGCAGACT	480
1484	QY	TCCTTGTGTGTACTTTTTCATCCAGAGTCTGTAGGGTTGCTTATGAAATCATGCAAAACAC	1543
481	Db	TCCTAGTGTGTACTTTTTCATCCAGAGTCTGTGAGTTGCTTATGAAATCATGCAAAACAC	540
1544	QY	TGCATCCTGATGCTCTGCAAAATTCATCTCTTTTGTAGATGACATCTACTATTTTGGAGGCC	1603
541	Db	TACGTCTGTAGTCTCTGCAAAATTCATCTCTTTTGTAGTATGATCACTACTATTTGGGGGCC	600
1604	QY	AAAATGCCCAACACAGATTGCAGTTTATCTCCACCAACTCGAACTAAAGAGGAAATTC	1663
601	Db	AGAAATGCCCAACATCAAATTTGCCATTTATGCTACACCAACCCGAACTGCAGATGAAATTC	660
1664	QY	CCATGNAACCTGGAGATATCATTTGGTGTGGCTGGAAACCAATTGGAATGGTTACTCTAAAG	1723
661	Db	CCATGGAACTGGAGATATCATTTGGTGTGGCTGGAAATCAATTGGGATGGCTATTTCCAAAG	720
1724	QY	GTGTCAACAGAAAATAGGAAAAACAGGCCCTGTACCCCTTCTCAAAAAGTCCGAGAGAAGA	1783
721	Db	GTGTCAACAGGAAATGGGGAAGGACNGCCCTATATCC-TCTCAAAAGTTCGAGAGAAGA	779
1784	QY	TAGAAACAGTCAAATACCCCTACATATCTCTGAAGCTGAAAAATAGAGAT	1831
780	Db	TAGAAACCGTCAAGTACCCACATATCTCTGGAGCTGAGAAATAAGCT	827

RESULT	13
BUSI31162	
LOCUS	
DEFINITION	BUSI31162 937 bp mRNA linear EST 12-SEP-2002 ACENGCOURT 10116686 NIH MGC 134 Mus musculus cDNA clone
ACCSSION	IMAGE:6508154 5', mRNA sequence.
VERSION	BUSI31162.1 GI:22820688
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	1 (bases 1 to 937) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE	Unpublished
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

```

Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUAM14073 row: 9 column: 03
High quality sequence stop: 751.
Location/Qualifiers
1. 937
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6508154"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dr. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."

```

RESULT 14
BX414472/c
LOCUS
DEFINITION

```

ACCESSION  BX414472
VERSION     BX414472.1  GI:30653877
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 757)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7499.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAN001ZF02_AN062_1&cluster=7499.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0BAN001ZF02_AN062_1.
Location/Qualifiers
1..757
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG005Y007"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT  191 a 160 c 167 g 238 t 1 others
ORIGIN
Query Match 31.6%; Score 634.6; DB 13; Length 757;
Best Local Similarity 90.1%; Pred. No. 3.3e-140;
Matches 679; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1098 ATACTTGATCCGTCCACACCTTGCTGGAAGGGAATAGAGAAACCCACCAAGACT 1157
DB 757 ATACTTGATCCGCCACACCTTGCTGCTAGAAAAGAAATAGAGAGCCNCCAGAGCT 698
QY 1158 TGGCTTCAAAATCCAGTTATTGGAGTCCATGTCTAGCGCACTGCACAAAGTGGGAACAGA 1217
DB 697 TGGCTTCAAAATCCAGTTATTGGAGTCCATGTCTAGCGCACTGCACAAAGTGGGAACAGA 638
QY 1218 AGCAGCTTCCATCCATGGAGATACATGTGTACAGTTCAGTTCAGTTCAGTTCCT 1277
DB 637 AGCTGCTTCCATCCATGGAGATACATGTGTACAGTTCAGTTCAGTTCAGTTCCT 578
QY 1278 CGAAGCGAGAAAGTGGATAAAGAGAGTGTATCTGGCCACTGATGACCCCTTCCTT 1337
DB 577 TCACGCGAGATGCAAGTGGACAAAAGAGTGTATTGGCCACAGATGACCCCTTCCTT 518
QY 1338 GTTAAAGGAGGCAAGACAAAGTACTCCAAATPATGAATTTATTAGTGATAACTCTATTTC 1397
DB 517 ATTTAAAGGAGGCAAGACAAAGTACTCCCAATTTATGAATTTATTAGTGATAACTCTATTTC 458
QY 1398 TTGGTCAGTGTGACTACACACCGATACACAGAAATTCATCTCGGCGGTGATCCCTGGA 1457
DB 457 CTGGTCAGTGTGACTGCAATTCGATACACAGAAATTCATCTCGGCGGTGATCCCTGGA 398
QY 1458 TATACATTTCTCTCCAGGCTGACTCTCTGTGTACTTTTTCATCCAGGCTCTGAG 1517
DB 397 TATACATTTCTCTCTCAGGAGACTTCTAGTGTGTACTTTTTCATCCAGGCTCTGAG 338
QY 1518 GGTGCTTATGAATCATGCACAACTGCATCTCGCTCTGCACAACTTCCATTTCTTT 1577

```

337 AGTTGCTTATGAAATTTATGAAACACTACATCTCTGATGCTCTCTCAAACTTCCATCTTT 278  
 1578 AGATGACATCTACTATTATTTGGAGGCCAAATATGCCCAACACAGATTCAGTTTATCTCTCA 1637  
 277 AGATGACATCTACTATTATTTGGGGCCAGATATGCCCAACATCAATTCGCAATTTATGCTCA 218  
 1638 CCAACCTCGAACTAAAGAGGAATCCCATCGAAACCTGGAGATATCATTTGGTGTGGCTGG 1697  
 217 CCAACCCGGAACCTGCAGATGAAATTCCTCATGGAACTCGAGATATCATTTGGTGTGGCTGG 158  
 1698 AAACCATTTGGAATGGTTACTCTAAAGCTGTCAACAGAAACTAGGAAACACAGCCCTGTA 1757  
 157 AAATCATTTGGGATGGCTATTCTAAAGGTGTCAACAGGAAATTTGGGAAGGACGGGCTATA 98  
 1758 CCCTTCTCTACAAAGCTCCGAGAGAGATAGAAACAGTCAAAATACCTCATCATATCTCTGAAGC 1817  
 97 TCCCTCTCTACAAAGTTCGAGAGAGATAGAAACGCTCAAGTACCCACATATCTCTGAGGC 38  
 1818 TGAATAATAGAGATGGAGTGTAGAGATTAACAA 1851  
 37 TGAGAAATAAAGCTCAGATCGAGAGAGATAAACGA 4

RESULT 15  
 CB989769 790 bp mRNA linear EST 01-MAY-2003  
 LOCUS AGENCOURT 13892856 NIH MGC 147 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30344717 5', mRNA sequence.  
 CB989769  
 ACCSSION CB989769  
 VERSION CB989769.1 GI:30284289  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 750)  
 NTH-MGC http://mgc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: csapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Stefan Hansson  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help  
 and advice from Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDAM381 row: h column: 06  
 High quality sequence stop: 619.  
 Location/Qualifiers  
 1..790  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30344717"  
 /tissue\_type="Human Placenta"  
 /lab\_host="DH10B Tona"  
 /clone\_lib="NIH\_MGC\_147"  
 /notes="Organ: placenta; Vector: pBluescriptR; Site\_1:  
 alt-XhoI; Site\_2: BamH; Oligo-dT primed using primer  
 5'-TTTTTTTTTTTTTTVN-3', size-selected for average  
 insert size 2.3 kb and normalized to ROT 5. This is a  
 primary library enriched for full-length clones and  
 constructed using the Cap-trapper method (Carninci, in  
 preparation). Library constructed by M. Brownstein  
 (NIMH/NHGRI, National Institutes of Health). Note: This is  
 a NIH MGC library."

BASE COUNT 244 a 157 c 202 g 186 t 1 others  
 ORIGIN

Query Match	31.68;	Score 634.6;	DB 14;	Length 790;
Best Local Similarity	93.0%;	Pred. No. 3.3e-140;		
Matches 675;	Conservative 0;	Mismatches 50;	Indels 1;	Gaps 1;
Qy	357	AGAAGAACAGCTTGTGTTAAGGCCAAAGAACAGATTGAAAAATTTACAAGAAACAAGCTAGGAA	416	
Db	27	AGAAGACAGCTTGTGTTAAGGCCAAAGAACAGATTGAAAAATTTACAAGAAACAAGCCAGAA	86	
Qy	417	TGATCTGGGAAAGGATCATGAAATCTTTAAGGAGGAGGATTGAAATGGAGCTTAAAGAGCT	476	
Db	87	TGGTCTGGGAAAGGATCATGAAATCTCTGAGGAGGAGGATTGAAATGGAGCTTAAAGAGCT	146	
Qy	477	CTGGTTTTTCTACAAAGTGAATTTGAAGAAATTTAAAGAAATTTAGAAGGAAACGAACCTCCA	536	
Db	147	CTGGTTTTTCTACAGAGTGAATTTGAAGAAATTTAAAGAACTTTAGAAGGAAATGAACCTCCA	206	
Qy	537	AAGACATGCAGATGAAATCTTTTGGATTTTAGGACATCATGAAAGGCTCTATCATGACAGA	596	
Db	207	AAGACATGCAGATGAAATCTTTTGGATTTTAGGACATCATGAAAGGCTCTATAATGACGGA	266	
Qy	597	TCTATACTACTCAGTCAACAGATGGAGCAGGTGAGTGGCGGGAAGAAAGAGCCAAAGA	656	
Db	267	TCTATACTACTCAGTCAACAGATGGAGCAGGTGATTGGCGGGAAGAAAGAGGCGCAAAGA	326	
Qy	657	TCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTGCAGAAATCCCAAGGACTGCAGCAA	716	
Db	327	TCTGACAGAACTGGTTCAGCGGAGAAATAACATATCTTCAGAAATCCCAAGGACTGCAGCAA	386	
Qy	717	AGCCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTATGGATGTCAACTCCATCA	776	
Db	387	AGCCAAAGCTGGTGTGTAATATCAACAAAGGCTGTGGCTATGGCTGTCAAGCTCCATCA	446	
Qy	777	TGTGGTTTACTGCTTCATGATTGCTTATGGCACCCAGCGAACACTCATCTTGGAACTCA	836	
Db	447	TGTGGTCTACTGCTTCATGATTGCATATGGCACCCAGCGAACACTCATCTTGGAAATCTCA	506	
Qy	837	GAAATGGCGCTATGCTACTGGAGGATGGGAGACTGTGTTTAGACCTCTGAAGTGAACATG	896	
Db	507	GAAATGGCGCTATGCTACTGGTGGATGGGAGACTGTATTTAGGCCTGTGAAGTGAACATG	566	
Qy	897	CACAGACAGGTCTGGCCTCTCCACTGGACACTGTGTCAGGTGAAGTGAAGGACAAAATGT	956	
Db	567	CACAGACAGATCTGGCATCTCCACTGGACACTGTGTCAGGTGAAGTGAAGGACAAAATGT	626	
Qy	957	TCAAGTGTTCAGCTCCCATTTGAGACAGCTCCATCTCGTCTCCTTACTTACCCTT	1016	
Db	627	TCAAGTGTTCAGCTTCCCATTTGAGACAGCTTTCATCCCGCTCCTCCACATTTACCCTT	686	
Qy	1017	GGCTGTACAGAAAGACCTTGCAGATCGACTCTGAGAGTCCAT -GGTGATCTCTGCAGTGT	1075	
Db	687	GGCTGTACAGAAAGACCTTGCAGATCGACTTGTACAGTGTGATGGGTGACCTGCAGTGN	746	
Qy	1076	GGTGGG 1081		
Db	747	TGGTGG 752		

Search completed: February 2, 2004, 18:36:52  
Job time : 4338.69 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 10:19:28 ; Search time 6342.17 Seconds  
(without alignments)  
11146.322 Million cell updates/sec

Title: US-09-971-773-2  
Perfect score: 1728  
Sequence: 1 atggggcagtgactgttc.....atctgaagctgaaaaatag 1728

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.on.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.hcgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1728	100.0	1728	6	BD168517	BD168517 Cells pro
2	1728	100.0	2976	10	BC010666	BC010666 Mus muscu
3	1723.2	99.7	2176	10	AB025198	AB025198 Mus muscu
4	1584	91.7	2008	6	BD168516	BD168516 Cells pro
5	1501.4	86.9	1759	6	AX662735	AX662735 Sequence
6	1501.4	86.9	1759	6	BD087745	BD087745 Plant cel
7	1501.4	86.9	2002	9	D89289	D89289 Homo sapien
8	1501.4	86.9	2100	6	EJ5725	EJ5725 Human mRNA
9	1501.4	86.9	2796	9	HSAS39535	AJ539535 Homo sapi
10	1501.4	86.9	2898	9	HSY17976	Y17976 Homo sapien
11	1501.4	86.9	2902	9	HSAS39536	AJ539536 Homo sapi
12	1501.4	86.9	2992	9	HSY17977	Y17977 Homo sapien
13	1501.4	86.9	3186	9	HSY17978	Y17978 Homo sapien
14	1501.4	86.9	3280	9	HSY17979	Y17979 Homo sapien
15	1501.4	86.9	3568	9	HSAS36055	AJ536055 Homo sapi
16	1501.4	86.9	3666	9	HSAS36053	AJ536053 Homo sapi
17	1501.4	86.9	3772	9	HSAS36054	AJ536054 Homo sapi
18	1501.4	86.9	4196	9	HSAS36056	AJ536056 Homo sapi
19	1496.6	86.6	2100	6	AR170077	AR170077 Sequence
20	1472.6	85.2	1728	4	AF247186	AF247186 Bos tauru
21	1469.4	85.0	1728	6	AR170074	AR170074 Sequence
22	1469.4	85.0	1728	6	E14720	E14720 Porcine mRN
23	1469.4	85.0	1836	4	D86723	D86723 Porcine mRN
24	1208	69.9	2682	9	BC025385	BC025385 Homo sapi
25	1088.2	63.0	2129	5	XLA514872	AJ514872 Xenopus l
26	941	54.5	1850	9	AF052088	AF052088 Homo sapi
27	931	53.9	979	6	BD168522	BD168522 Cells pro
28	894.2	51.7	979	6	BD168521	BD168521 Cells pro
29	876.2	50.7	1017	6	E43925	E43925 Antibody re
30	755.4	43.7	1772	9	AB049740	AB049740 Homo sapi
31	604.6	35.0	699	6	E43926	E43926 Antibody re
32	401.6	23.2	1262	9	HSAS14325	AJ514325 Homo sapi
33	401.6	23.2	1396	9	HSAS14324	AJ514324 Homo sapi
34	378.8	21.9	1860	3	AF411264	AF411264 Drosophil
35	378.8	21.9	2764	3	AY051451	AY051451 Drosophil
36	357.2	20.7	2860	3	AK113018	AK113018 Ciona int
37	322.6	18.7	844	11	BV033231	BV033231 S212P6677
38	322.6	18.7	171073	2	AC132342	AC132342 Mus muscu
39	316.6	18.3	110000	2	AC129041	Continuation (2 of
40	316.6	18.3	110000	2	AC129041	Continuation (3 of
41	316.6	18.3	150709	2	AC134206	AC134206 Rattus no
42	316.6	18.3	237954	2	AC128532	AC128532 Rattus no
43	296.4	17.2	394	6	BD071198	BD071198 Secreted
44	271	15.7	1147	9	AB049828	AB049828 Homo sapi
45	271	15.7	186158	9	CNS018P0	AL109847 Human chr

ALIGNMENTS

RESULT 1  
BD168517  
LOCUS BD168517 1728 bp DNA linear PAT 17-JAN-2003  
DEFINITION Cells producing antibody composition.  
ACCESSION BD168517  
VERSION BD168517.1 GI:27874329  
KEYWORDS WO 0231140-A/2.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1728)  
AUTHORS Kanda,Y., Sato,M., Nakamura,K., Uchida,K., Shinkawa,T., Yamane,N.,  
Hosaka,E., Yamasaki,M. and Hanai,N.  
TITLE Cells producing antibody composition

JOURNAL	Patent: WO 0231140-A 2 18-APR-2002;	
COMMENT	OS KYOWA HAKKO KOGYO CO LTD	
	PN Mus musculus (mouse)	
	PD WO 0231140-A/2	
	PF 18-APR-2002	
	PP 05-OCT-2001 WO 2001JP008804	
	PR 06-OCT-2000 JP 00P 308526	
	PI YUTAKA KANDA, MITSUO SATO, KAZUYASU NAKAMURA, KAZUHISA UCHIDA, PI	
	TOYOHIDE SHINKAWA, NAOKO YAMANE, EMI HOSAKA, MOTOO YAMASAKI, NOBUO	
	PI HANAI	
	PC C12N5/10, C12P21/08, C07K16/00, A01K67/00, A61K39/395, C12N9/00, PC	
	C12N15/52,	
	PC GOIN33/53	
	CC Cells producing antibody composition	
	PH Key Location/Qualifiers	
	PT source 1..1728	
FEATURES	Location/Qualifiers	
source	1..1728	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:10090"	
BASE COUNT	529 a 365 c 400 g 434 t	
ORIGIN		
Query Match	100.0%; Score 1728; DB 6; Length 1728;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1728;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGCGGGCATGACCTGGTTCCTGGGCTTGGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60	661 GGTGTCACTCCATCAGCTGGTCTACTGTTTTCATGATTGCTTATGGCACCCAGCAACA 720
DB	1 ATGCGGGCATGACCTGGTTCCTGGGCTTGGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60	661 GGTGTCACTCCATCAGCTGGTCTACTGTTTTCATGATTGCTTATGGCACCCAGCAACA 720
QY	61 TTGTTATTTTATATAGTGGTTCATTTGTTCCAGATAATGACCCCTGATCAGTCCAGC 120	721 CTCACTTTGGAAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGGAGACTGTGTTTGA 780
DB	61 TTGTTATTTTATATAGTGGTTCATTTGTTCCAGATAATGACCCCTGATCAGTCCAGC 120	721 CTCACTTTGGAAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGGAGACTGTGTTTGA 780
QY	121 AGAGAACTCTCCAAGATTCTTCAAGCTTGAACCTTAAACAGCAAAATGAAGACTTG 180	781 CCTGTAAGTGGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGTGTCAGTGAA 840
DB	121 AGAGAACTCTCCAAGATTCTTCAAGCTTGAACCTTAAACAGCAAAATGAAGACTTG 180	781 CCTGTAAGTGGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGTGTCAGTGAA 840
QY	181 AGCGAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCATTTGACCGGGACAGCTACA 240	841 GTAAATGACAAAACATTTCAAGTGGTCCAGCTCCCCATTTGTAGACAGGCTCCATCCTCGG 900
DB	181 AGCGAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCATTTGACCGGGACAGCTACA 240	841 GTAAATGACAAAACATTTCAAGTGGTCCAGCTCCCCATTTGTAGACAGGCTCCATCCTCGG 900
QY	241 GGAAGAGTCCGTTTGTAGAGAACAGCTTTGTAAGGCCCAAGAACAGATTTGAAATTTAC 300	901 CCTCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960
DB	241 GGAAGAGTCCGTTTGTAGAGAACAGCTTTGTAAGGCCCAAGAACAGATTTGAAATTTAC 300	901 CCTCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960
QY	301 AAGAAACAAGCTAGAAATGGTCTGGGAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360	961 GGTGACCTGTCAGTGGTGGTGTCCAGATTTGTCCAGTTTGTCAAAATACTTGATTGCTCCACA 1020
DB	301 AAGAAACAAGCTAGAAATGGTCTGGGAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360	961 GGTGACCTGTCAGTGGTGGTGTCCAGATTTGTCCAGTTTGTCAAAATACTTGATTGCTCCACA 1020
QY	361 AATGAGCTAAAGAGCTGTGTTTCTTCAAGCGAATCTGAAGAAATTTAAAGCATTTA 420	1021 TGGCTGGAAAAAGGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATCCTCAGTTATT 1080
DB	361 AATGAGCTAAAGAGCTGTGTTTCTTCAAGCGAATCTGAAGAAATTTAAAGCATTTA 420	1021 TGGCTGGAAAAAGGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATCCTCAGTTATT 1080
QY	421 GAAGGAATGAACTCCAAGACATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480	1081 GGAGTCCATGTCCAGCGCACAGAAAGTGGGAAACAGAAAGCCTTCCACCCATCGAG 1140
DB	421 GAAGGAATGAACTCCAAGACATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480	1081 GGAGTCCATGTCCAGCGCACAGAAAGTGGGAAACAGAAAGCCTTCCACCCATCGAG 1140
QY	481 AGGTCTATCATGACAGATCTACTACCTCAGTCAAAACAGATGGAGCGGGATTGGCGT 540	1141 GAGTACATGGTACAGCTTGAAGAAATTTTCAGCTTCTCCAGCGCAGAAATGCAAGTGGAT 1200
DB	481 AGGTCTATCATGACAGATCTACTACCTCAGTCAAAACAGATGGAGCGGGATTGGCGT 540	1141 GAGTACATGGTACAGCTTGAAGAAATTTTCAGCTTCTCCAGCGCAGAAATGCAAGTGGAT 1200
QY	541 GAAAAGAGGCCAAAGATCTGACAGAGTGGTTCAGCGGAGAAATTAATATCTCCAGAA 600	1201 AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTGTAAAGGAGGCAAGACAAAG 1260
DB	541 GAAAAGAGGCCAAAGATCTGACAGAGTGGTTCAGCGGAGAAATTAATATCTCCAGAA 600	1201 AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTGTAAAGGAGGCAAGACAAAG 1260
QY	601 CCTAAGGACTGCAGCAAGCCAGGAGCTGGTGTGTAACATCAATAAAGGCTGTGCTAT 660	1261 TACTCCATTAATGAATTTATTAGTGAATCTATTTCTGTCAGCTGGACTACACAAT 1320
DB	601 CCTAAGGACTGCAGCAAGCCAGGAGCTGGTGTGTAACATCAATAAAGGCTGTGCTAT 660	1261 TACTCCATTAATGAATTTATTAGTGAATCTATTTCTGTCAGCTGGACTACACAAT 1320
		1321 CGGTACACAGAAAAATTCACATTCGGGGTGTGATCTCTGGATATACACTTTCTCTCACAGGCT 1380
		1321 CGGTACACAGAAAAATTCACATTCGGGGTGTGATCTCTGGATATACACTTTCTCTCACAGGCT 1380
		1381 GACTTTCTAGTGTGTACTTTTTCATCCAGGCTGTGTCGGGTTGCTTATGAATCATGCA 1440
		1381 GACTTTCTAGTGTGTACTTTTTCATCCAGGCTGTGTCGGGTTGCTTATGAATCATGCA 1440
		1441 ACCCTGCATCTGATGCTCTCGCAATTCCTCACTTTTGGATGACATCTACTATTTCGA 1500
		1441 ACCCTGCATCTGATGCTCTCGCAATTCCTCACTTTTGGATGACATCTACTATTTCGA 1500
		1501 GGCCAAAATGCCCAATCAGATTCGTTTATCTCAAAAACCTCGAACTGAAGAGGAA 1560
		1501 GGCCAAAATGCCCAATCAGATTCGTTTATCTCAAAAACCTCGAACTGAAGAGGAA 1560
		1561 ATTCCAATGGAACTGGAGATATCATTTGGTGTGGTGGAAACCATTTGGATGGTTATTCT 1620
		1561 ATTCCAATGGAACTGGAGATATCATTTGGTGTGGTGGAAACCATTTGGATGGTTATTCT 1620
		1621 AAGGTATCAACAGAAACTTTGGAAACAGGCTTATATGCTCTCTCAAAAGTCCGAGAG 1680
		1621 AAGGTATCAACAGAAACTTTGGAAACAGGCTTATATGCTCTCTCAAAAGTCCGAGAG 1680
		1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAAATAG 1728
		1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAAATAG 1728

```
RESULT 2
BC010666          2976 bp      mRNA      linear      ROD 16-APR-2003
LOCUS             Mus musculus fucosyltransferase 8, mRNA (cdna clone MGC:11418
DEFINITION        IMAGE:3594582), complete cds.
ACCESSION         BC010666
VERSION           BC010666.1 GI:14715012
KEYWORDS          MGC.
SOURCE            Mus musculus (house mouse)
ORGANISM          Mus musculus
REFERENCE          1 (bases 1 to 2976)
AUTHORS           Strausberg, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
                  Klausner, R.D., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
                  Altschul, S.F., Zeeberg, B., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
                  Hopkins, R.F., Jordan, H., Maruina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
                  Diatchenko, L., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
                  Stappleton, M., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
                  Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
                  Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
                  Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
                  McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
                  Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
                  Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
                  Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
                  Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
                  Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
                  Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
                  Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
                  Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
                  Generation and initial analysis of more than 15,000 full-length
                  human and mouse cDNA sequences
                  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 2976)
Strausberg, R.
Direct Submission
Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: c9apbs@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LILNL at: http://image.llnl.gov
Series: IRAK Plate: 16 Row: 1 Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8393370.
Location/Qualifiers
1..2976
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/tissue="MGC:11418 IMAGE:3594582"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
```

```
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..2976
/gene="Fut8"
/db_xref="LocusID:53618"
/db_xref="MGI:1858901"
584..2311
/codon_start=1
/product="fucosyltransferase 8"
/protein_id="AAHI0666.1"
/db_xref="GI:14715013"
/db_xref="LocusID:53618"
/translaton="MRAWTGSWRWIMLIFAWGTLLFYIGHVLVDNDHPHSHRELS
KIALKRLKQONEDLRMAESLRIPGPIDQGTATGRVRVLEBQLVKAKEQIENYKK
KARNGLGKDEHLLRRRTENGAKELWFFLOSELKKLKHLEGNELORHAEILLDLGHHE
RSIMTDLVYLSOTDAGDWEKEAKDITELVORITVLYONPKDCSKARKLVNKNKG
GYGQOLHNVIVCFMIAVTQTRLLESQNRHRYATGGWETVFRPVSCTDRSGLSGTGH
WSGVNDHNIQVVELPVLVDLSHPPPLPLAVPDLADLRLLRVHGDPAVWVSVQVYK
LIRPQWLEKEIEBATPKGPHVIGVHVRRTDKVGEAAAFHPIEYMHVVEHFLQ
LARMQVDKRVLATDDPTLLKEAKTKYSNYEFISONSISWSAGLHNRVYTESLFGV
ILDTHFLQADFLVCTFSSOVRVAYEIMOTLHPDASANPHSLDDIYVFGQNAHQI
AVYPHKRTEEEIPWEFDIIGVAGNHWGYSKGINRKLKGLTGLYPSYKVKREKIEIVK
YPTYPEAKK"
BASE COUNT      909 a      626 c      660 g      781 t
ORIGIN
Query Match      100.0%; Score 1728; DB 10; Length 2976;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCGGGGCGATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60
DB 584 ATCGGGGCGATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 643
QY 61 TTGTTATTTATATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 120
DB 644 TTGTTATTTATATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 703
QY 121 AGAGAACTCTCCAGATTCTTGCAAGCTTCAACCGCTTAAACAGCAAGAAATGAAGACTTG 180
DB 704 AGAGAACTCTCCAGATTCTTGCAAGCTTCAACCGCTTAAACAGCAAGAAATGAAGACTTG 763
QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGGCCCCCATTTGACGGGGACAGCTACA 240
DB 764 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGGCCCCCATTTGACGGGGACAGCTACA 823
QY 241 GGAAGAGTCCGTGTTTAGAAGAAACAGCTTCTTTAAGGCCAAAGAACAGATTGAAAATTAC 300
DB 824 GGAAGAGTCCGTGTTTAGAAGAAACAGCTTCTTTAAGGCCAAAGAACAGATTGAAAATTAC 883
QY 301 AAGAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360
DB 884 AAGAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 943
QY 361 AATGAGGCTAAAGAGTCTGCTGTTTTTTCTACAAAGCGAACTCAAGAAATTTAAAGCATTTA 420
DB 944 AATGAGGCTAAAGAGTCTGCTGTTTTTTCTACAAAGCGAACTCAAGAAATTTAAAGCATTTA 1003
QY 421 GAAGAAATGAATCTCCAAAGACATCGAGATGAAATCTTTTGGATTTAGGACCATCATGAA 480
DB 1004 GAAGAAATGAATCTCCAAAGACATCGAGATGAAATCTTTTGGATTTAGGACCATCATGAA 1063
QY 481 AGGTCTATCATGACAGATCTATCTACTCTAGTCAACAGATGGAGCAGGGGATTTGGCGT 540
DB 1064 AGGTCTATCATGACAGATCTATCTACTCTAGTCAACAGATGGAGCAGGGGATTTGGCGT 1123
QY 541 GAAAAAGAGGCCAAAGATCTCACAGAGCTGTGTCAGCGGAGAAATAACATATCTCCAGAAAT 600
DB 1124 GAAAAAGAGGCCAAAGATCTCACAGAGCTGTGTCAGCGGAGAAATAACATATCTCCAGAAAT 1183
QY 601 CCTAAGGACTCGAGCAAGAGCCAGGAAGCTGGTGTGTATCAATCAATAAAGGCTGGCTAT 660
DB 1184 CCTAAGGACTCGAGCAAGAGCCAGGAAGCTGGTGTGTATCAATCAATAAAGGCTGGCTAT 1243
```







QY 1 ATGCGGCAATGACCTGGTTCCTGGGCTTGGATATATGCTCAATCTTTTGGCTCGGGGACC 60  
DB |||||  
QY 100 ATGCGGCAATGACCTGGTTCCTGGGCTTGGATATATGCTCAATCTTTTGGCTCGGGGACC 159  
DB |||||  
QY 61 TTGTTATTTATATAGTGGTCAATTTGGTTCGAGATATGACCCCTGATCACTCCAGC 120  
DB |||||  
QY 160 TTATTTGTTTATATAGTGGTCAATTTGGTTCGAGATATGACCCCTGATCACTCCAGC 219  
DB |||||  
QY 121 AGAGAACTCTCCAGATCTTCCAAAGCTTGAACCTTTAAACACAGCAAAATGAAGACTTG 180  
DB |||||  
QY 220 AGAGAACTCTCCAGATCTTCCAAAGCTTGAACCTTTAAACACAGCAAAATGAAGACTTG 279  
DB |||||  
QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCTCATTTGACCGGGGACAGCTACA 240  
DB |||||  
QY 280 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCTCATTTGATCAGGGGACAGCTACA 339  
DB |||||  
QY 241 GGAAGAGTCCGTTTATAGAGAACAGCTTGTTAAGGCCAAAGAACAGATTTGAAAATTTAC 300  
DB |||||  
QY 340 GGAAGAGTCCGTTTATAGAGAACAGCTTGTTAAGGCCAAAGAACAGATTTGAAAATTTAC 399  
DB |||||  
QY 301 AAGAAACAAAGCTTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAGAGAGGAGATTGAA 360  
DB |||||  
QY 400 AAGAAACAAAGCTTAGAAATGATCTGGGGAAGGATCATGAAATCTTAAGAGAGGAGATTGAA 459  
DB |||||  
QY 361 AATGAGCTTAAAGAGCTCTGGTCTTCTCAAAAGCGAACTGAAGAAATTAAGCAATTTA 420  
DB |||||  
QY 460 AATGAGCTTAAAGAGCTCTGGTCTTCTCAAAAGCGAACTGAAGAAATTAAGCAATTTA 519  
DB |||||  
QY 421 GAAGGAATGAATCTCAAGACATCGAGATGAATATCTTTGGATTTAGGACACCATGAA 480  
DB |||||  
QY 520 GAAGGAATGAATCTCAAGACATCGAGATGAATATCTTTGGATTTAGGACATCATGAA 579  
DB |||||  
QY 481 AGGTCTATCATGACAGATCTACTACTCTCAGTCAAAAGGATGAGAGAGGAGTTGGCGT 540  
DB |||||  
QY 580 AGGTCTATCATGACAGATCTACTACTCTCAGTCAAAAGGATGAGAGAGGAGTTGGCGG 639  
DB |||||  
QY 541 GAAAGAGGCGCAAGAGCTGACAGAGCTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 600  
DB |||||  
QY 640 GAAAGAGGCGCAAGAGCTGACAGAGCTGGTCCAGCGGAGAAATTAACATATCTCCAGAA 699  
DB |||||  
QY 601 CCTAAGGAGTGCAGAAAGCCAGGAGCTGGTGTGTATCATCAATTAAGGCTGTGGCTAT 660  
DB |||||  
QY 700 CCCAAGGAGTGCAGAAAGCCAGAAAGCTGGTATGTAATATCAACAAAGGCTGTGGCTAT 759  
DB |||||  
QY 661 GTTCTGCAACTCCATCACTGCTGCTACTGTTTCTATGATTGCTTATGCGCACCCAGCGAACA 720  
DB |||||  
QY 760 GAGTGTCAACTCCATCATGCTGGTGTCTGCTTCTATGATTGCTTATGCGCACCCAGCGAACA 819  
DB |||||  
QY 721 CTCACTTTGGAACTCTCAGAAATGGCGCTATGCTACTGGTGGATGGAGACTGTGTTTGA 780  
DB |||||  
QY 820 CTCACTTTGGAACTCTCAGAAATGGCGCTATGCTACTGGAGGATGGAGACTGTGTTTGA 879  
DB |||||  
QY 781 CCTGTAAGTGAACATGTATGACAGAGATCTGGCGCTCTCCACTGGACACTGGTCAAGTGAA 840  
DB |||||  
QY 880 CCTGTAAGTGAACATGTATGACAGAGATCTGGCGCTCTCCACTGGACACTGGTCAAGTGAA 939  
DB |||||  
QY 841 GTAATGACAAACATTTCAAGTGTCTGAGCTCCCAATTTGATAGACAGCTCCATCTCCGG 900  
DB |||||  
QY 940 GTGAAGGACAAATGTTTCAAGTGTCTGAGCTCCCAATTTGATAGACAGCTCCATCTCGT 999  
DB |||||  
QY 901 CCTCTTACTTACCACTGGCTGTTCCAGAAAGCTTTGACAGACCGCACTCTTAAGAGTCCAT 960  
DB |||||  
QY 1000 CCTCTTACTTACCTTGGCTGTACAGAGACCTTGCAGATCGACTCTCGAGAGTCCAT 1059  
DB |||||  
QY 961 GGTGACCTCGAGTGTGGTGTCCAGATTTGTCAAATATCTGATTCGTCCCAACCT 1020  
DB |||||  
QY 1060 GGTGATCTCGAGTGTGGTGTCCAGATTTGTCAAATATCTGATTCGTCCCAACCT 1119  
DB |||||  
QY 1021 TGGCTGGAAGAGGAAATAGAGAACCCACCAAGAGCTTGGCTTCAACATCTCAGTTATT 1080  
DB |||||  
QY 1120 TGGCTGGAAGAGGAAATAGAGAACCCACCAAGAGCTTGGCTTCAACATCTCAGTTATT 1179  
DB |||||

QY 1081 GGAGTCCATGTCAGACGACAGACAAAGTGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB |||||  
QY 1180 GGAGTCCATGTCAGACGACAGACAAAGTGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239  
DB |||||  
QY 1141 GAGTACATGGTACACGCTTGAAGAACATTTTTCAGCTTCTCGCACGACAGAAATGAAGTGGAT 1200  
DB |||||  
QY 1240 GAATACATGGTACACGCTTGAAGAACATTTTTCAGCTTCTCGAACGACAGAAATGAAGTGGAT 1299  
DB |||||  
QY 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTACTTTGTTTAAAGAGGAGGAGAGAGAGAGAG 1260  
DB |||||  
QY 1300 AAAAAAGAGTGTATCTGGCCACTGATGACCTTTCTTTTAAAGAGGAGGAGAGAGAGAGAGAG 1359  
DB |||||  
QY 1361 TACTCCAAATTAAGAAATTAAGTAACTCTATTTCTTGGTCACTGAGTCACTACCAAT 1320  
DB |||||  
QY 1360 TACTCCAAATTAAGAAATTAAGTAACTCTATTTCTTGGTCACTGAGTCACTACCAAC 1419  
DB |||||  
QY 1321 CGGTACACAGAAATTCACCTTCGGGGTGTGATCTCGATATACACTTTCTCTCACAGGCT 1380  
DB |||||  
QY 1420 CGATACAGAAATTCACCTTCGGGGTGTGATCTCGATATACACTTTCTCTCACAGGCT 1479  
DB |||||  
QY 1381 GACTTCTTAGTGTACTTTTTTCAATCCAGGTCTGTCGGGTGCTTATGAAATCATGCAA 1440  
DB |||||  
QY 1480 GACTTCTTAGTGTACTTTTTTCAATCCAGGTCTGTCAGGTGCTTATGAAATCATGCAA 1539  
DB |||||  
QY 1441 ACCGTGATCTGATGCTCTCGAACTTCATTTCTTTGGATGACATCTACTATTTTGA 1500  
DB |||||  
QY 1540 AACTGATCTGATGCTCTGCAAACTTCATTTCTTTAGATGACATCTACTATTTTGA 1599  
DB |||||  
QY 1501 GGCCAAATGCCACAAATCAGATTGCTTTATCTCACAACCTCGAACTGGAAGAGAA 1560  
DB |||||  
QY 1600 GGCCAAATGCCACAAACAGATTGCAATTTATCTCACAACCTCGAACTGGAAGAGAA 1659  
DB |||||  
QY 1561 ATTCAAATGGAACCTCGAGATATCATTTGGTGTGCTGGAACCAATGGGATGGTTATCT 1620  
DB |||||  
QY 1660 ATCCCATGNAACCTCGAGATATCATTTGGTGTGCTGGAACCAATGGATGGTTATCT 1719  
DB |||||  
QY 1621 AAAGGTATCAACAGAAACCTTGGAAACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1680  
DB |||||  
QY 1720 AAAGGTGTCAACAGAAACCTAGGAAACAGGCTGTACCTCTCTCAAAAGTCCGAGAG 1779  
DB |||||  
QY 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAATATAG 1728  
DB |||||  
QY 1780 AAGATAGAAACAGTCAAGTATCCCTACATATCTCTGAAGCTGAAATATAG 1827  
DB |||||

RESULT 5  
AX662735 LOCUS AX662735 1759 bp DNA linear PAT 22-MAR-2003  
DEFINITION Sequence 1 from Patent WO02070672.  
ACCESSION AX662735  
VERSION AX662735.1 GI:29163324

KEYWORDS  
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

Fujiyama, K., Seki, T. and Taniguchi, N.

Plant cell having animal-type sugar chain adding function

TITLE Patent: WO 02070672-A 1 12-SEP-2002;

JOURNAL Fujiyama, Kazuhito (JP); Seki, Tatsuji (JP)

FEATURES

Location/Qualifiers

1..1759

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

17..1744

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAD79985.1"

/db\_xref="GI:29163325"

/translation="MRPWTGSRWIMLILFANGTLLFVIGSHLVRDNDHPDHSRELS

KILARLRKQNEEDLRMAESLRIPGPDQGPAGRVRLVLEQLVKAKEQIENYKK

QTRNLGKHDBILRRRIENGAKELWFFLQSELKKLKNLEGNELQRLHAFDEFLDLGHHE  
RSIMTDLVLSQTAGDWREKADLTTELVRITTYLONPKDCSKAKLVNINKGC  
GYCOLHHVVYCFMAYCTORTLILESONWRYATGWETVFPVSECTDRSGISTGH  
WSBEVKDNQVQVPLPVIDSLHPPPLPLAVPEDLADRLVRVHGDPAVWVSQFVKY  
LIRPOPEKEIBEATKKGKPGVIGVHVRRTDKVGEAAFPHEIETWVHEHFPQL  
LARMQVKKRVLATDDPSLLKKAATKYPNBFISONSISWSAGLHNRVTENSLRGV  
ILDHFLSQADFLVCTSSQVRYAYEIMOTLHPDASANFSLDDIYFQGNANHQI  
AIYAHQRTADEIPMEPGDIIIGVAGNHWGYSKGVNRKLGRTGLYPSYKVRKLEIYVK  
YPTYPEARK"  
BASE COUNT 533 a 367 c 418 g 441 t  
ORIGIN

Query Match 86.9%; Score 1501.4; DB 6; Length 1759;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGGGGGATGGAGTGGTCTCGGGTGGGATTGCTCATCTCTTTTGGCTGGGGACC 60  
DB 17 ATGGGGCATGGAGTGGTCTCGGGTGGGATTGCTCATCTCTTTTGGCTGGGGACC 76  
QY 61 TTGTTATTTTATATAGGTGGTCAATTTGGTTCGAGATAATGACCCTGTGATCACTCCAGC 120  
DB 77 TTGCTGTTTTATATAGGTGGTCACTTGGTACGAGATAATGACCCTGTGATCACTCTAGC 136  
QY 121 AGAGAACTCTCAAGATTCTTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
DB 137 CGAGAACTGTCCAAGATTCTGGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 196  
QY 181 AGGGAAATGGCTGAGTCTCTCGAATACAGAAAGGCCCATGTGACGGGGACAGCTACA 240  
DB 197 AGGGAAATGGCGGAATCTCTCGGATACAGAAAGGCCCATGTGATGACGGGGCCAGCTATA 256  
QY 241 GGAAGAGTCCGTGTTTGAAGACAGCTTGTGAAGCCAAAGACAGATTTGAATTTAC 300  
DB 257 GGAAGAGTACGGTTTTGAAGAGAGAGCTTGTGAAGCCAAAGACAGATTTGAATTTAC 316  
QY 301 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360  
DB 317 AAGAAACAGACCAAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 376  
QY 361 AATGAGCTAAAGAGCTCTGTTTTTCTCAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
DB 377 AATGAGCTAAAGAGCTCTGTTTTTCTCAAGCGAACTGAAGAAATTTAAAGCACTTA 436  
QY 421 GAAGAAATGAATCTCAAGACATGAGATGAATTTCTTTGGATTAGGACACCATGAA 480  
DB 437 GAAGAAATGAATCTCAAGACATGAGATGAATTTCTTTGGATTAGGACATCATGAA 496  
QY 481 AGGTCTATCATGACAGATCTATACCTCAGTCAACAGATGGAGCAGGGGATTGGCGT 540  
DB 497 AGGTCTAATGACGGATCTATACCTCAGTCAACAGATGGAGCAGGTGATTGGCGG 556  
QY 541 GAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATACATATCTCCAGAA 600  
DB 557 GAAAGAGGCGCAAGATCTGACAGAACTGGTTCAGCGGAGAAATACATATCTCCAGAA 616  
QY 601 CCTAAGGATCGACAAAGCCAGGAGCTGGTGTGATACATCAATTAAGGCTGTGGCTAT 660  
DB 617 CCCAAGGATCGACAAAGCCAAAGCTGGTGTGATATCAACAAAGGCTGTGGCTAT 676  
QY 661 GGTGTCAACTCCATCAGTGGTCTACTGTTTCATGATTGCTTATGGCACCCAGCGAACA 720  
DB 677 GGCTGTGAGTCCATCATGTGGTCTACTGTTTCATGATTGATATGGCACCCAGCGAACA 736  
QY 721 CTCAFTTTGGAATCTCAGAAATGGCGCTATGCTACTGTGGATGGAGACTGTGTTTGA 780  
DB 737 CTCAFTTTGGAATCTCAGAAATGGCGCTATGCTACTGTGGATGGAGACTGTATTTAGG 796  
QY 781 CCTGTAAGTGAGACATGTACAGAGATCTGGCTCTCCACTGGACACTGTGTCAGTGAA 840  
DB 797 CCTGTAAGTGAGACATGTACAGAGATCTGGCTCTCCACTGGACACTGTGTCAGTGAA 856  
QY 841 GTAAATGACAAAACATTTCAAGTGGTCGAGCTCCCCCATTTGTAGACAGCTCCATCCTCGG 900

DB 857 GTGAAGGACAAAAATGTTCAAGTGGTGGAGTTCCTCCATTTGTAGACAGTCTTTCATCCCGGT 916  
QY 901 CCTCTTACTTACCACATGGCTGTTTCCAGAGACCTTCAGACCGCCTCTTAAGAGTCCAT 960  
DB 917 CCTCATATTTACCCCTTGGCTGTACCAAGAACCTCGCAGATCGACTTGTACGAGTGT 976  
QY 961 GGTGACCCCTGCAAGTGTGGTGGTCTCCAGTTTGTCAAAATCTTGAATTCGTCCCAACCT 1020  
DB 977 GGTGACCCCTGCAAGTGTGGTGGTCTCAGTTTGTCAAAATCTTGAATTCGTCCCAACCT 1036  
QY 1021 TGGCTGGAAGAAAGAAATAGAAAGACCAACCAAGAGCTTGGCTTCAAAACATCCAGTTAT 1080  
DB 1037 TGGCTAGAAAAGAAATAGAAAGACCAACCAAGAGCTTGGCTTCAAAACATCCAGTTAT 1096  
QY 1081 GGAGTCCATGTTCAGACGCACAGACAAAAGTGGAAAGAGAGCTTCCACCCCATCCAG 1140  
DB 1097 GGAGTCCATGTTCAGACGCACAGACAAAAGTGGAAAGAGAGCTTCCATCCCATGAA 1156  
QY 1141 GAGTACATGTTACAGTGGTGTGATGATCTCTTTCAGGCTTCTCGCACGACAGATCGAGT 1200  
DB 1157 GAGTACATGTTGATGTTGAGAACATTTTCAGCTTCTTCGACGACAGATCGAGTGGAC 1216  
QY 1201 AAAAAAGAGTATATCTGGGTACTGATGATCTCTTGTGTTTAAAGAGGCAAAAGACAAAG 1260  
DB 1217 AAAAAAGAGTATTTGGCCACAGATGACCCCTTCTTTTAAAGAGGCAAAAGACAAAG 1276  
QY 1261 TACTCCAAATGATGATTTATGATGATACTCTATTTCTTGGTGGTGGTGGTGGTGGTGGT 1320  
DB 1277 TACCCCAATATGAAATTTATGATGATACTCTATTTCTTGGTGGTGGTGGTGGTGGTGGT 1336  
QY 1321 CGGTACACAGAAAATTCATCTCGGGGTGTGATCTCGATATACATCTTCTCTCAGAGCT 1380  
DB 1337 CGATACACAGAAAATTCATCTCGGGGTGTGATCTCGATATACATCTTCTCTCAGAGCA 1396  
QY 1381 GACTTCTAGTGTCTATCTTTTCAATCCAGGTCTGTGGGTGTGTTTGAATTAAGAAATCATGCA 1440  
DB 1397 GACTTCTAGTGTCTATCTTTTCAATCCAGGTCTGTGGGTGTGTTTGAATTAAGAAATCATGCA 1456  
QY 1441 ACCCTGATCTGATGCTCTCGAACTTCTCAATTTTGGTGGTGGTGGTGGTGGTGGTGGT 1500  
DB 1457 ACATGATCTGATGCTCTCGAACTTCTCAATTTTGGTGGTGGTGGTGGTGGTGGTGGT 1516  
QY 1501 GGCCAAATGCCCAATCAGATTGCTTGTATCTCACAACCTCGAACTCGAAGAGGAA 1560  
DB 1517 GGCCAAATGCCCAATCAGATTGCTTGTATCTCACAACCTCGAAGAGGAA 1576  
QY 1561 ATTCCAAATGGAACTGAGATATCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1620  
DB 1577 ATTCCCAATGGAACTGAGATATCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1636  
QY 1621 AAAGGTATCAACAGAAATTTGGAAACAGGCTTATCTCCTCAAAAGTCCGAGAG 1680  
DB 1637 AAAGGTGTCAACAGAAATTTGGAAAGAGCGGCTTATCTCCTCAAAAGTTCGAGAG 1696  
QY 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATA 1727  
DB 1697 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATA 1743

## RESULT 6

BD087745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BD087745 1759 bp DNA linear PAT 27-AUG-2002  
Plant cells having function to add animal type sugar chain.

BD087745

BD087745.1 GI:22633355

JP 2001333787-A/1.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1759)

Taniguchi, N., Seki, T. and Fujiyama, K.

TITLE Plant cells having function to add animal type sugar chain  
JOURNAL Patent: JP 2001333787-A 1 04-DEC-2001;  
NAOYUKI TANIGUCHI, TATSUJI SEKI, KAZUHIRO FUJIYAMA  
COMMENT OS Homo sapiens (human)  
PN JP 2001333787-A/1  
PD 04-DEC-2001  
PF 06-MAR-2001 JP 2001062704  
PI NAOYUKI TANIGUCHI, TATSUJI SEKI, KAZUHIRO FUJIYAMA PC  
C12N15/00, A01H5/00, C12N5/10, C12N9/10, C12P19/00, C12P21/02, PC  
C12N15/00,  
PC C12N5/00  
CC Plant cells having function to add animal type sugar chain FH  
Key Location/Qualifiers  
FT CDS (17)..(1744).  
FEATURES  
source  
BASE COUNT 533 a 367 c 418 g 441 t  
ORIGIN  
Query Match 86.9%; Score 1501.4; DB 6; Length 1759;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 1 ATGCGGGCATGACGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 60  
DB 17 ATGCGGGCATGACGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 76  
QY 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCACTTCATCTCCAGC 120  
DB 77 TTGCTGTTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCACTTCATCTTAGC 136  
QY 121 AGAAGACTCTCAAGATCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTG 180  
DB 137 CGAGAACTGTCCAAGATCTGCGAAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCT 196  
QY 181 AGGCGAATGGCTGAGTCTCTCGAATACCAAGAGCCCATTCAGAGGCGGACAGCTACA 240  
DB 197 AGCGAATGGCGCAATCTCTCGAATACCAAGAGCCCATTCAGAGGCGGACAGCTACA 256  
QY 241 GGAAGAGTCCGTGTTTGAAGAACAGCTTGTTGAAGGCCAAAGAACAGACTGAAATTCAC 300  
DB 257 GGAAGAGTACGGTTTGAAGAACAGCTTGTTGAAGGCCAAAGAACAGACTGAAATTCAC 316  
QY 301 AAGAAACAGCTAGAAATGGTCTGGGAAGGATCATGAATCTTAAGAGGAGGATTCGA 360  
DB 317 AAGAAACAGACAGAAATGGTCTGGGAAGGATCATGAATCTTAAGAGGAGGATTCGA 376  
QY 361 AATGAGCTTAAAGAGCTCTGGTTTCTTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
DB 377 AATGAGCTTAAAGAGCTCTGGTTTCTTCAAGAGTGAATGAAGAAATTTAAAGAACTTA 436  
QY 421 GAAGGAAATGAATCTCAAGAACATGCAGATGAATTTCTTTTGGATTTAGGACACCATGAA 480  
DB 437 GAAGGAAATGAATCTCAAGAACATGCAGATGAATTTCTTTTGGATTTAGGACATCATGAA 496  
QY 481 AGGTCTATCAGACAGATCTATACCTCAGTCAACAGATGAGAGGAGGATTTGGCGT 540  
DB 497 AGGTCTATATGACGGATCTATACCTCAGTCAACAGATGAGAGGAGGATTTGGCGG 556  
QY 541 GAAAGAGGCGCAAGATCTGACAGAGTGGTTCAGAGGAGGATTAACATATCTCCAGAA 600  
DB 557 GAAAGAGGCGCAAGATCTGACAGATGGTTCAGAGGAGGATTAACATATCTCCAGAA 616  
QY 601 CTTAAGGACTGCAGCAAGCCAGGAGTGGTGTGTAAATCAATCAATTAAGGCTGTGGCTAT 660  
DB 617 CCAAGGACTGCAGCAAGCCAGGAGTGGTGTGTAAATCAATCAATTAAGGCTGTGGCTAT 676  
QY 661 GGTGTCACTCATCAGTGGTCTACTGTTTCATGATTTGCTATGCGACCCAGCGAACA 720  
DB 677 GGCTGTGAGCTCCATCATGTGGTCTACTGCTTCATGATTTGATGCGACCCAGCGAACA 736

QY 721 CTCACTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGCTGATGCGGAGACTGTGTTAGA 780  
DB 737 CTCACTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGCTGATGCGGAGACTGTATTTAGG 796  
QY 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGACACTGCTGAGGTGAA 840  
DB 797 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGACACTGCTGAGGTGAA 856  
QY 841 GTAAATGACAAAAACATTTCAAGTGGTGGCTCCCATTTGTAGACAGCTTCATCCCGCT 900  
DB 857 GTGAGGACAAAAATGTTCAAGTGGTGGCTCCCATTTGTAGACAGCTTCATCCCGCT 916  
QY 901 CCTCTTTACTTACCATCTGGCTTTCCAGAGACCTTGGAGACCACTCTTAAGAGTCCAT 960  
DB 917 CCTCCATATTTACCTTTGGCTTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCAT 976  
QY 961 GGTGACCTCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1020  
DB 977 GGTGACCTCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1036  
QY 1021 TGGCTGGAAGAGAAATAGAAAGCAACCAAGAGCTTTGGCTTTCAAAACATCCAGTTATT 1080  
DB 1037 TGGCTGGAAGAGAAATAGAAAGCAACCAAGAGCTTTGGCTTTCAAAACATCCAGTTATT 1096  
QY 1081 GGAGTCCATGTACAGACGACAGCAAAAGTGGGAAACAGAAAGCGCTTCCACCCCATCGAG 1140  
DB 1097 GGAGTCCATGTACAGACGACAGCAAAAGTGGGAAACAGAAAGCGCTTCCACCCCATCGAA 1156  
QY 1141 GAGTACATGTACAGCTTTGAAGACATTTTCAGCTTTTGGCAGCAGAAATGCAAGTGGAC 1200  
DB 1157 GAGTACATGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1216  
QY 1201 AAAAAAGAGTATATCTGGCTTACTGATGATCTTCTTCTTAAAGAGGCAAGACAAAG 1260  
DB 1217 AAAAAAGAGTATTTTGGCCACAGATGACCTTTCTTTTAAAGAGGCAAGACAAAG 1276  
QY 1261 TACTTCCATATGAATTTATTTAGTAACTCTATATTTCTTGGTGGTGGTGGTGGTGGTGGT 1320  
DB 1277 TACCCCAATTTATGAATTTATTTAGTAACTCTTATTTCTTGGTGGTGGTGGTGGTGGT 1336  
QY 1321 CGGTACACAGAAATTTCACTTCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1380  
DB 1337 CGATACACAGAAATTTCACTTCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1396  
QY 1381 GACTTCTAGTGTGATCTTTTCACTCCAGCTCTGTGGGTGGTGGTGGTGGTGGTGGTGGT 1440  
DB 1397 GACTTCTAGTGTGATCTTTTCACTCCAGCTCTGTGGGTGGTGGTGGTGGTGGTGGTGG 1456  
QY 1441 ACCCTGATCTGATGCTCTGCAACTTCCATTTCTTTGGTGGTGGTGGTGGTGGTGGTGG 1500  
DB 1457 ACCTATATCTGATGCTCTGCAACTTCCATTTCTTTAGTAACTCTACTATTTTGGG 1516  
QY 1501 GGCCAAATGGCCCAATCAGATTTGCTTTTATCTCTCAAACTCGAAGCTGAGAGGAA 1560  
DB 1517 GGCCAGATGGCCCAATCAGATTTGCTTTATGCTCACCACCCCGCAACTGAGATGAA 1576  
QY 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1620  
DB 1577 ATTCCCATGGNACCTGGAGATATCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1636  
QY 1621 AAAGGTATCAACAGAAAACTTTGGAAAAACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1680  
DB 1637 AAAGGTGTCAACAGAAAACTTTGGAAAAAGGACGGGCTTATATCTCTCTCAAAAGTCCGAGAG 1696  
QY 1681 AAGATAGAAACAGTCAAGTATCCCATATCTCTCAAGCTGAATAA 1727  
DB 1697 AAGATAGAAACGGTCAAGTATCCCATATCTCTGAGGCTGAGAAATA 1743  
RESULT 7  
DB9289  
LOCUS DB9289 2002 bp mRNA linear PRI 07-FEB-1999

DEFINITION Homo sapiens mRNA for N-Acetyl-beta-D-glucosaminide, complete cds.  
 D89289.1 GI:2055306  
 VERSION N-Acetyl-beta-D-glucosaminide; GDP-L-Fuc; alpha 1-6  
 KEYWORDS Fucosyltransferase; alphas-1-6 FucT.

SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 YANAGIDANI, S., Uozumi, N., Ihara, Y., Miyoshi, E., Yamaguchi, N. and  
 Taniguchi, N.

TITLE Purification and cDNA cloning of  
 GDP-L-Fuc:N-Acetyl-beta-D-glucosaminide:alpha-1-6 fucosyltransferase  
 (alpha-1-6 FucT) from human gastric cancer MKN45 cells  
 J. Biochem. 121 (3), 626-632 (1997)

JOURNAL 97279058  
 MEDLINE 9133635  
 PUBMED  
 REFERENCE  
 AUTHORS Yanagidani, S.

TITLE Purification and cDNA cloning of  
 GDP-L-Fuc:N-Acetyl-beta-D-glucosaminide:alpha 1-6  
 Fucosyltransferase (alpha 1-6 FucT) from human stomach carcinoma  
 MKN45 cells

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 2002)  
 AUTHORS Taniguchi, N.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-NOV-1996) Naoyuki Taniguchi, Osaka University Medical  
 School, Department of Biochemistry, Yamadaoka 2-2, Suita, Osaka  
 565, Japan (E-mail: proftani@biochem.med.osaka-u.ac.jp,  
 Tel:81-6-879-3420, Fax:81-6-879-3429)

FEATURES  
 source  
 1..2002  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 143..1870  
 /note="GDP-L-Fuc; alpha 1-6 Fucosyltransferase; alpha-1-6  
 FucT"  
 /codon\_start=1  
 /product="N-Acetyl-beta-D-glucosaminide"  
 /protein\_id="BAA19764.1"  
 /db\_xref="GI:2055307"

/translation="MRPTGSRWIMLILFANGTLLFYIGHLVRDNDHPDHSRELS  
 KILAKLERKQONEDLRMAESLRIPEDIPDGPAGIRVRVLEQLVKAQGIENYKK  
 QTRNGLRKHDELRRIENGAKELFFLOSELKLLKNLEGNELQRAHDFLLDLGHE  
 RSIIMDLVYLSOTDGDGWRKEAKDLTELVRITLYLQNPDKCSKAKKLVNINCKG  
 GYGCCLHVVYCFMAYGTORTLILESQNRVATGCVETFPVSETCTDRSGISTGH  
 WSGEVKDNQVQVELPIVDSLHPRPYPLAVPEDLADLRVHGDPAVWVSOFVKY  
 LIRPQVLEKEIEATKGLGFKHPVIGVHVRRTDKVTEAFPHPIEYMWVHEHFQL  
 LARRMQDKRVYLDTPSLLEAKTKYPNYEFISDMSISWSAGLHNRITYENSURGV  
 ILDIHFLSQADPLVCTFSQVCRVAYEIMOTLHPDASANPHSLDDIYFFGQNAHQI  
 AIYAHQPRTADEIPMEPGDIIIGVAGNWDGYSKGVNRKLGRTGLYPSYKVKREIETVK  
 YPTYPEAK"

BASE COUNT 620 a 415 c 467 g 500 t  
 ORIGIN  
 Query Match 86.9%; Score 1501.4; DB 9; Length 2002;  
 Best Local Similarity 91.8%; Pred. No. 0;  
 Matches 1596; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGCGGCGATGACTGGTTCCTGGCGTTGGATTATGCTCATCTTTTGGCTCGGGGACC 60  
 Db 143 ATGCGGCGATGACTGGTTCCTGGCGTTGGATTATGCTCATCTTTTGGCTCGGGGACC 202  
 QY 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCAGATATGACACCCCTGATCATCTCCAGC 120  
 Db 203 TTGCTGTTTTATATAGTGGTTCATTTGGTTCAGATATGACACCCCTGATCATCTTAGC 262  
 QY 121 AGAGAACTCTCAAGATCTTGCAGAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
 Db 263 CGAGAACTCTCAAGATCTTGCAGAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 322

181 AGCGAATGGCTGACTCTCTCCGAATACCAGAGGCCCCCAATTGTACAGGGGACAGCTACA 240  
 Db 323 AGCGAATGGCCGAATCTCTCCGATACCAGAGGCCCTATTGTATCAGGGGCCAGCTATA 382  
 QY 241 GGAAGAGTCCGTGTTTGAAGAAACAGCTTGTGAAGCCCAAGAACAGATGAAAATTAC 300  
 Db 383 GGAAGAGTACCGCTTTTGAAGAGCAGCTTGTGAAGCCCAAGAACAGATGAAAATTAC 442  
 QY 301 AAGAAAACAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360  
 Db 443 AAGAAAACAGCCAGAAATGGTCTGGGAAGGATCATGAAATCTTGAAGAGGAGATTGAA 502  
 QY 361 AATGAGAGTAAAGAGCTCTGGTCTTCTCAAAAGCCGAACCTGAAGAAATTAAGACATTTA 420  
 Db 503 AATGAGAGTAAAGAGCTCTGGTCTTCTCAAGAGTGAATTTGAAGAAATTAAGAACTTA 562  
 QY 421 GAAGAAATGAATCTCAAGAGCATGCAGATGAATAATCTTTGGATTTAGGACACCATGAA 480  
 Db 563 GAAGAAATGAATCTCAAGAGCATGCAGATGAATAATCTTTGGATTTAGGACATCATGAA 622  
 QY 481 AGGTCTATCATGACAGATCTTACTCTCAGTCAAAACAGATGAGCAGGGGATTTGGCGT 540  
 Db 623 AGGTCTATATGACGATCTTACTCTCAGTCAAGAGCATGCAGATGAGCAGGGGATTTGGCGG 682  
 QY 541 GAAAAGAGGCCAAATCTGACAGAGCTGGTCCAGGGGAGAAATAATATCTCCAGAAAT 600  
 Db 683 GAAAAGAGGCCAAATCTGACAGAGCTGGTCCAGGGGAGAAATAATATCTTCAGAAAT 742  
 QY 601 CCTAAGAGTGCACGAAAGCCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
 Db 743 CCCAAGAGTGCACGAAAGCCAAAGCTGGTGTGTAAATATCAACAAAGGCTGTGGCTAT 802  
 QY 661 GGTGTCAATCCATCAGCTGCT 720  
 Db 803 GGCTGTGAGCTCCATATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 862  
 QY 721 CT 780  
 Db 863 CT 922  
 QY 781 CCT 840  
 Db 923 CCT 982  
 QY 841 GTAAATGACAAAACATTCAGTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
 Db 983 GTGAAGGACAAAATGTTCAAGTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1042  
 QY 901 CCT 960  
 Db 1043 CCT 1102  
 QY 961 GGTGACCTCTGAGTGTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
 Db 1103 GGTGACCTCTGAGTGTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1162  
 QY 1021 TGGCTGGAAGGAAAATGAAGAGCCACCAAGAGCTTGGCTTCAACATCCAGCTTAT 1080  
 Db 1163 TGGCTGGAAGGAAAATGAAGAGCCACCAAGAGCTTGGCTTCAACATCCAGCTTAT 1222  
 QY 1081 GGAGTCCATGTACAGCAGACAGAAAAGTGGGAAGCAGAGCTTCCACCCCATCCAG 1140  
 Db 1223 GGAGTCCATGTACAGCAGACAGAAAAGTGGGAAGCAGAGCTTCCATCCCATCCAG 1282  
 QY 1141 GAGTACATGGTACAGCTTGAAGAAACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAT 1200  
 Db 1283 GAGTACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1342  
 QY 1201 AAAAAAGAGTATATCTGGCTATGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260  
 Db 1343 AAAAAAGAGTATATCTGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1402

1261 TACTCAATATGAATTTATATAGTATGATAAATCTTATTTCTGTCTAGCTGGACTACACAAT 1320  
1403 TACCCCAATATGAATTTATATAGTATGATAAATCTTATTTCTGTCTAGCTGGACTGACAAAT 1462  
1321 CGGTACACAGAAATTCACCTCGGGGTGATGATCCTGGATATACATTTCTCTCAGAGCT 1380  
1463 CGATACACAGAAATTCACCTCGGTGAGTATGATCCTGGATATACATTTCTCTCAGGCA 1522  
1381 GACTTCTAGTGTACTTTTTTATCCAGAGTCTGTGGGTTGCTTATGAATCATGCAA 1440  
1523 GACTTCTAGTGTACTTTTTTATCCAGAGTCTGTGGGTTGCTTATGAATCATGCAA 1582  
1441 ACCCTGATCTGATGCTCTCGAACTTCCATTTCTTTGGATGACATCTACTATTTTGA 1500  
1583 ACATACATCTCTGATGCTCTCGAACTTCCATTTCTTTAGATGACATCTACTATTTTGG 1642  
1501 GGCCTAATGCCACAAATCAGATTTCTGTTTATCTCTCAAACTTCAAGTGAAGAA 1560  
1643 GGCCTAATGCCACAAATCAGATTTCTGTTTATCTCTCAAACTTCAAGTGAAGAA 1702  
1561 ATTCCAAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGTTTATCT 1620  
1703 ATTCCAAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGTTTATCT 1762  
1621 AAAGGTATCAACAGAAACTTGGAAACAGAGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
1763 AAAGGTATCAACAGAAACTTGGAAAGAGCGGCTTATATCCCTCTCAAAAGTCCGAGAG 1822  
1681 AGATAGAAACAGTCAAGTATCCCATATCTCTGAGCTGAAATA 1727  
1823 AAAGTAGAAACGTTCAAGTATCCCATATCTCTGAGCTGAAATA 1869

RESULT 8

E15725 LOCUS 2100 bp DNA linear PAT 28-JUL-1999  
DEFINITION Human mRNA for alpha-1,6-fucosyltransferase, complete cds.  
ACCESSION E15725  
VERSION E15725.1 GI:5710408  
KEYWORDS JP 1998084975-A/1  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TANIGUCHI N., UOZUMI H. and YANAGIYA S.  
ALPHA-1,6-FUCOSYLTRANSFERASE GENE DERIVED FROM HUMAN  
Patent: JP 1998084975-A 1 07-APR-1998;  
TOYOBO CO LTD  
OS Homo sapiens (human)  
PN JP 1998084975-A/1  
PD 07-APR-1998  
PP 17-JUN-1997 JP 1997159692  
PI 22-JUL-1996 JP 96P 192260  
PR TANIGUCHI NAOYUKI, UOZUMI HISAFUMI, YANAGIYA SHUSAKU PC  
C12N15/09, C07H21/04, C12N1/21, C12N9/10, (C12N15/09, C12N1/91), PC  
(C12N1/21,  
PC C12R1/19), (C12N9/10, C12R1/19);  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1. .2100  
FT FT /organism='Homo sapiens'  
FT FT /cell\_type='stomach cancer cell' FT  
FT /cell\_line='MKN45'  
FT CDS 198. .1925  
FT Location/Qualifiers  
FT 1. .2100  
FT /organism='Homo sapiens'  
FT /mol\_type='genomic DNA'  
FT /db\_xref='taxon:9606'

BASE COUNT 651 a 436 c 489 g 524 t  
ORIGIN  
Query Match  
Best Local Similarity 86.9%; Score 1501.4; DB 6; Length 2100;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
Qy 1 ATGCGGCATGGACTGGTCTCTGCGGTGGATATGCTCATTTCTTTTGCCTGGGGACC 60  
Db 198 ATGCGGCATGGACTGGTCTCTGCGGTGGATATGCTCATTTCTTTTGCCTGGGGACC 257  
Qy 61 TTGTTATTTATATAGTGTCTTTGTTTCGAGATTAATGACCACTTATGACCTCATCCAGC 120  
Db 258 TTGCTGTTTTATATAGTGTGCTTCTGATACGAGATTAATGACCTTATGATCCTCTAGC 317  
Qy 121 AGAGAACTCTCCAAAGATTTCTGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
Db 318 CGAGAACTGTCCAAGATTTCTGGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 377  
Qy 181 AGCGAATGGCTGAGTCTCTCCGAATACGAAAGGCCCCATTTGACAGGGGACAGCTACA 240  
Db 378 AGCGAATGGCGGAATCTCTCCGATACGAAAGGCCCCATTTGATCAGGGGCCAGCTATA 437  
Qy 241 GGAAGAGTCCGTGTTTGTAGAGAACAGCTTTTAAAGCCCAAGAACAGATTTGAAATTTAC 300  
Db 438 GGAAGAGTACCGCTTTTGTAGAGAACAGCTTTTAAAGCCCAAGAACAGATTTGAAATTTAC 497  
Qy 301 AAGAAACAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAAGAGAGGATTTGAA 360  
Db 498 AAGAAACAGCAAGAAATGGTCTGGGAAGGATCATGAAATCTTGAAGAGGAGGATTTGAA 557  
Qy 361 AATGAGCTAAAGAGCTCTGTTTTTTCTACAAAGCGAACTGAAAGAAATTAAGCATTTA 420  
Db 558 AATGAGCTAAAGAGCTCTGTTTTTTCTACAGAGTGAAATTTGAAGAAATTAAGCACTTA 617  
Qy 421 GAAGAAATGAATCTCCAAAGCATGAGATGAAATTTCTTTTGGATTTAGGACCATGAA 480  
Db 618 GAAGAAATGAATCTCCAAAGCATGAGATGAAATTTCTTTTGGATTTAGGACCATGAA 677  
Qy 481 AGGTCTATATGACAGATCTATCTACTCTCAGTCAAAACAGATGGAGAGGAGTTGGCGT 540  
Db 678 AGGTCTATATGACCGATCTATCTACTCTCAGTCAAGACAGATGGAGAGGAGTTGGCGG 737  
Qy 541 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATTAACATATCTCCAGAT 600  
Db 738 GAAAGAGGCGCAAGATCTGACAGAACTGGTTCAGCGGAGAAATTAACATATCTCTCAGAT 797  
Qy 601 CCTAAGAGCTGCAGCAAAAGCCAGGAAGCTGGTGTGAACATCATTAAGGCTGTGGCTAT 660  
Db 798 CCTAAGAGCTGCAGCAAAAGCCAGGAAGCTGGTGTGAATATCAACAAAGGCTGTGGCTAT 857  
Qy 661 GGTGTCAATCTCCATCACGTGGTCTACTGTTTCATGATTTGCTTATGGCACCAGCGAACA 720  
Db 858 GGTGTCAATCTCCATCACGTGGTCTACTGTTTCATGATTTGCTTATGGCACCAGCGAACA 917  
Qy 721 CTCACTTTGGAATCTCAGAAATTTGGCGCTATGCTACTCTGTTGATGGAGAGCTGTGTTAGA 780  
Db 918 CTCACTTTGGAATCTCAGAAATTTGGCGCTATGCTACTCTGTTGATGGAGAGCTGTGTTAGG 977  
Qy 781 CCTGTAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGGTGAA 840  
Db 978 CCTGTAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGGTGAA 1037  
Qy 841 GTAAATGACAAAAACATTTCAAGTGGTGGAGCTCCCATTTGTAGACAGCTTCATCTCCG 900  
Db 1038 GTGAGAGCAAAATTTTCAAGTGGTGGAGCTTCATTTGTAGACAGCTTCATCTCCGCT 1097  
Qy 901 CCTCTTTTACCACTGGCTGTTTCCAGAGAGACTTTCAGACCGAGCTCTTAAGAGTCCAT 960  
Db 1098 CCTCCATATTTACCTTTGGCTGTACAGAGAGACTTTCAGACAGCTTCATCTCCGAGTGCAT 1157  
Qy 961 GGTGACCTCTGAGTGGTGGTGTCCAGTTTGTCAATACTTGTCTGCTCCACACCT 1020

```
Db 1158 GGTGACCTCGAGTGTGGTGGTGTCTCAGTTTGTCAAATATCTTGATCCGCCACAGCCT 1217
Qy 1021 TGGCTGGAAGAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTATT 1080
Db 1218 TGGCTAGAAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTATT 1277
Qy 1081 GGAGTCCATGTGACAGCGACAGCAAAAGTGGGAAGAGAGCTTCCACCCCATCGAG 1140
Db 1278 GGAGTCCATGTGACAGCGACAGCAAAAGTGGGAAGAGAGCTTCCATCCCATCGAA 1337
Qy 1141 GAGTACATGTTGACAGTTGAAGACATTTTTCAGCTTCTCGACGACAGATCAAGTGGAT 1200
Db 1338 GAGTACATGTTGACAGTTGAAGACATTTTTCAGCTTCTCGACGACAGATCAAGTGGAC 1397
Qy 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTCTACTTTTAAAGGAGGCAAGACAAAG 1260
Db 1398 AAAAAAGAGTATATCTGGCCACAGATGACCTTCTTTTAAAGGAGGCAAGACAAAG 1457
Qy 1261 TACTCCAAATATGAATTTATTAGTGAATCTTATTTCTTGGTCAGCTGGACTACACAAT 1320
Db 1458 TACCCCAATATGAATTTATTAGTGAATCTTATTTCTTGGTCAGCTGGACTGCACAAT 1517
Qy 1321 CGGTACACAGAAAATTCATCTCGGGGTGATCTCGGATATACATTTTCTCTCACAGCT 1380
Db 1518 CGATACACAGAAAATTCATCTCGGGAGTGATCTCGGATATACATTTTCTCTCACAGCA 1577
Qy 1381 GACTTCTAGTGTGTACTTTTTCATCCAGGTGTGTCGGTTCGCTTATGAATCATGCAA 1440
Db 1578 GACTTCTAGTGTGTACTTTTTCATCCAGGTGTGTCGGTTCGCTTATGAATCATGCAA 1637
Qy 1441 ACCCTGCACTCTGATGCTCTCGCAATCTTCCATCTTTTGGATGACATCTACTATTTTGA 1500
Db 1638 ACCTACATCTGATGCTCTCGCAATCTTCCATCTTTTAGATGACATCTACTATTTTGG 1697
Qy 1501 GCGCAAAATGCCCAATCAGATTCCTGTTTATCTCTCAAAACCTCGAATCGAAGAGAA 1560
Db 1698 GCGCAAAATGCCCAATCAGATTCCTGTTTATCTCTCAAAACCTCGAATCGAAGAGAA 1757
Qy 1561 ATTCCAATGGAACTGGAGATATCATTTGTTGGTGGCAACCATTTGGATGGTTATTCT 1620
Db 1758 ATTCCAATGGAACTGGAGATATCATTTGTTGGTGGCAACCATTTGGATGGTTATTCT 1817
Qy 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680
Db 1818 AAAGGTATCAACAGGAAATTCGGAAGAGCGGCTATATCTCTCTCAAAAGTTCGAGAG 1877
Qy 1681 AAGATAGAAAAGTCAAGTATCCCAATATCTCTGAAGCTGAAAATA 1727
Db 1878 AAGATAGAAAAGTCAAGTATCCCAATATCTCTGAAGCTGAAAATA 1924
```

```
RESULT 9
HSA539535 HSA539535 2796 bp mRNA linear PRI 03-FEB-2003
LOCUS Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), clone
DEFINITION CD-L1.
ACCESSION AJ539535
VERSION AJ539535.1 GI:28200807
KEYWORDS alpha6-fucosyltransferase; FUT8 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mollicone, R., Michalski, J.C., Bauvy, C., Calileau-Thomas, A.,
Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and
Oriol, R.
TITLE Splice variants of alpha6-fucosyltransferase are expressed early in
human embryogenesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2796)
AUTHORS Oriol, R.
TITLE Direct Submission
```

## JOURNAL

## FEATURES

## source

```
Submitted (31-JAN-2003) Oriol R., U504, Inserm, 16 Av. Paul
Vaillant-Couturier, 94807, FRANCE
Location/Qualifiers
1..2796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23.2"
/clone="CD-L1"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
/country="France:Paris"
1..2796
/gene="FUT8"
326..2053
/gene="FUT8"
/EC_number="2.4.1.68"
/function="adds fucose in alphanol,6 linkage to the internal
GlcNAc of the core chitobiose"
/codon_start=1
/evidence="experimental"
/product="alpha6-fucosyltransferase"
/protein_id="CAD62444.1"
/db_xref="GI:28200808"
/translation="MRPWTGSRWIMLIFAWGTLLFYIGHVLVRDNDHPDHSRELS
KILAKLERLQONEDILRRMAESLRIPEGPIQDOGPAIGRVRLVLEBOLVKAKEQIENYKK
OTRNLGKDEHLLRRRIENGAKELWFLQSELKKLKNLEGNELORHAEFLDLIGHHE
RSIMTDIYLSQTDGADGWEKAKDITELVORITYLQNPDKDSKAKLVLNKGK
GSGQLHHVYCFMIAVGTQKLTILLESQNRWYATGWETFRPVSCTCTDKSGISTGH
WSEEDKQNVQVVELPIVLSLHPRPYPLVLAPEDLADRLVRVHGDPAVMMVSQFVK
LIRPQWLEKEIEBEATKKLGHPIGVHVRRTDKVGEAFAFPIESYMHVENSILGV
LARMQVDRKRVYLATDDPSLLKEAKTKYPNYEFISONSISWSAGLHNRYPENSLRGV
ILDHFEFSQADFLVCTPSSQCVAVEMOTLHPDASANFSLDDIYVFGQNAHNOI
AIYAHQRIADEIPMEPEDIIGVAGNHWGDSKGVNRKLGRTGLYPSYKVKREKJETVK
YPTYPEAEK"
polyA_signal 2771
polyA_site 2777
/gene="FUT8"
BASE COUNT 874 a 552 c 605 g 765 t
ORIGIN
Query Match 86.9%; Score 1501.4; DB 9; Length 2796;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy 1 ATCGCGCATGGACTGTCTCTCGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60
Db 326 ATCGCGCATGGACTGTCTCTCGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGACC 385
Qy 61 TTGTTATTTATATAGTGTGTCATTTGGTTCGAGATAATGACCACTGATCATCTCCAGC 120
Db 386 TTGCTGTTTTATATAGTGTGTCATTTGGTACGAGATAATGACCACTGATCATCTAGC 445
Qy 121 AGAGAACTCTCCAAGATTTCTTCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180
Db 446 CGAGAACTGTCCAAGATTTCTTCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 505
Qy 181 AGCGCAATGGCTGAGTCTCTCCGAATACCAAGAGCCCAATTTGACCGGGGACAGCTACA 240
Db 506 AGCGCAATGGCGGAATCTCTCCGATACCAAGAGCCCAATTTGATCAGGGGCCAGCTATA 565
Qy 241 GGAAGAGTCCGTGTTTGTAGAAAGACAGCTTGTAAAGCCAAAGAACAGATGAAATTTAC 300
Db 566 GGAAGAGTACGCGTTTGTAGAAAGACAGCTTGTAAAGCCAAAGAACAGATGAAATTTAC 625
Qy 301 AAGAAACAGCTAGAAATGTTCTCGGGAAGGATCATGAAATCTTAAAGAGAGGATTGAA 360
Db 626 AAGAAACAGCTAGAAATGTTCTCGGGAAGGATCATGAAATCTTAAAGAGAGGATTGAA 685
Qy 361 AATGGAGCTAAGAGCTCTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 420
```



Db 686 AATGAGCTAAGAGCTCTGGTTTCTACAGAGTGAATTGAAGAAATTAAGAACTTA 745  
Qy 421 GAAGGAAATGAATCCAAAGACATGCAGATGAATTTCTTTGGATTTAGGACACCATGAA 480  
Db 746 GAAGGAAATGAATCCAAAGACATGCAGATGAATTTCTTTGGATTTAGGACATCATGAA 805  
Qy 481 AGGCTATCATGACAGATCTATACCTCAGTCAACAGATGGAGAGGGGATTTGGCT 540  
Db 806 AGGCTATAATGACGGATCTATACCTCAGTCAACAGATGGAGAGGGGATTTGGCGG 865  
Qy 541 GAAGAAGAGCCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAA 600  
Db 866 GAAGAAGAGCCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAA 925  
Qy 601 CTTAAGGAGCTGACGAAAGCCAGGAGCTGGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
Db 926 CCCAAGGAGCTGACGAAAGCCCAAGAGCTGGTGTAAACATCAATAAAGGCTGTGGCTAT 985  
Qy 661 GGTCTCAATCCATCACTGGTCTACTGTTTCATGATTTATGGCAACCCAGCGAAACA 720  
Db 986 GGTCTCAATCCATCACTGGTCTACTGTTTCATGATTTATGGCAACCCAGCGAAACA 1045  
Qy 721 CTCATCTTGGATCTCAGAAATGGCGCTATGCTACTGCTGGTGGAGACTGTATTTAGG 780  
Db 1046 CTCATCTTGGATCTCAGAAATGGCGCTATGCTACTGCTGGTGGAGACTGTATTTAGG 1105  
Qy 781 CTTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCCACTGGACACTGGTCAAGTGAA 840  
Db 1106 CTTGTAAGTGAGACATGTACAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGTGAA 1165  
Qy 841 GTAATGACAAACCAATCAAGTGTGAGCTCCCATTTAGACAGAGCTCCATCTCGG 900  
Db 1166 GTGAAGGACAAACCAATCAAGTGTGAGCTCCCATTTAGACAGAGCTCCATCTCGG 1225  
Qy 901 CCTCTTACTTACACTGGCTGTTCCAGAGAGCTTGCAGACCGACTCCTAAGAGTCCAT 960  
Db 1226 CCTCTTACTTACACTGGCTGTTCCAGAGAGCTTGCAGACCGACTCCTAAGAGTCCAT 1285  
Qy 961 GGTGACCTCGAGTGTGGTGGTGTCCAGAGTGTTCCTCAATCTGATTCGTCACCAAGCT 1020  
Db 1286 GGTGACCTCGAGTGTGGTGGTGTTCCTCAATCTGATTCGTCACCAAGCT 1345  
Qy 1021 TGGCTGGAAGAAATGAAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1080  
Db 1346 TGGCTGGAAGAAATGAAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTAT 1405  
Qy 1081 GGAGTCCATGTGACAGCGACACAGAAAGTGGGACAGAGCGCTTCCACCCCATCGAG 1140  
Db 1406 GGAGTCCATGTGACAGCGACACAGAAAGTGGGACAGAGCGCTTCCATCCCATGAA 1465  
Qy 1141 GAGTACATGTTGACAGTGTGAAGAACATTTTTCAGCTTCTCGACGACAGAAATGCAAGTGGAT 1200  
Db 1466 GAGTACATGTTGACAGTGTGAAGAACATTTTTCAGCTTCTCGACGACAGAAATGCAAGTGGAT 1525  
Qy 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTCTTTGTTAAAGGAGGCAAGCAAG 1260  
Db 1526 AAAAAAGAGTATATTTGGCCACAGATGACCCCTCTTTTAAAGGAGGCAAGCAAG 1585  
Qy 1261 TACTCCAATTATGAATTTATGATGATTAATCTATTTCTTGGTCACTGAGCTACACAT 1320  
Db 1586 TACCCCAATTATGAATTTATGATGATTAATCTATTTCTTGGTCACTGAGCTGACCAAT 1645  
Qy 1321 CGGTACACAGAAATTCACCTCGGGGTGTGATCCCTGGATATACATTTCTCTCACAGGCT 1380  
Db 1646 CGATACACAGAAATTCACCTCGGGGTGTGATCCCTGGATATACATTTCTCTCACAGGCT 1705  
Qy 1381 GACTTTCTAGTGTGATCTTTTTCATCCAGAGTCTGTCGGGTGCTTATGAATATCATGAA 1440  
Db 1706 GACTTTCTAGTGTGATCTTTTTCATCCAGAGTCTGTCGGGTGCTTATGAATATCATGAA 1765  
Qy 1441 ACCCTGATCCTGATGCTGCGAACTTCATCTTTGGATGATCATCTACTATTTTGGG 1500  
Db 1766 ACCTGATCCTGATGCTGCGAACTTCATCTTTGGATGATCATCTACTATTTTGGG 1825

Qy 1501 GGCCAAATGCCACAATCAGATTGCTGTTTATCTCACAACCTCGAATCGAAGAGAA 1560  
Db 1826 GGCCAGAATGCCACAATCAGATTGCTGTTTATCTCACAACCTCGAATCGAAGAGAA 1885  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCACTGGATGGTTATTTCT 1620  
Db 1886 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGCTGGAATCAATTTGGATGGCTATTTCT 1945  
Qy 1621 AAAGGTATCAACAGAAACTTTGGAAAAACAGGCTTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 1946 AAAGGTATCAACAGAAACTTTGGAAAGGAGCGGCTATATCCCTCTCAAAAGTTCGAGAG 2005  
Qy 1681 AAGTAGAAGACAGTCAAGATATCCACATATCTCTGAAAGTGAATAAATA 1727  
Db 2006 AAGTAGAAGACAGTCAAGTATCCCAATATCTCTGAGGCTGAGAAATA 2052

RESULT 10  
LOCUS HSY17976 2898 bp mRNA linear PRI 17-SEP-1998  
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase, transcript B1.  
ACCESSION Y17976  
VERSION Y17976.1 GI:3451262  
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Cailleau, A., Balanzino, L., Candellier, J.J., Oriol, R. and Mollicone, R.  
TITLE Differential splice variants of human FUT8 embryonic cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2898)  
AUTHORS Cailleau, A.L.V.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V. Couturier, 94807 Villejuif Cedex, FRANCE  
COMMENT Related entry: D89289.  
FEATURES  
source  
1..2898  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q23"  
/clone="FUT8 transcript B1"  
/tissue\_type="whole embryo"  
/dev\_stage="40-70 days"  
1..2898  
/gene="FUT8"  
1..101  
/gene="FUT8"  
/note="B"  
102..200  
/gene="FUT8"  
/note="C"  
201..2156  
/gene="FUT8"  
/note="D"  
429..2156  
/gene="FUT8"  
/EC\_number="2.4.1.68"  
/note="allele 1"  
/codon\_start=1  
/product="glycoprotein 6-alpha-L-fucosyltransferase"  
/protein\_id="CAA76985.1"  
/db\_xref="GI:3451263"  
/db\_xref="SPTREMBL:O00235"  
/translation="MRPWTGSRWMLILFAWGTLFLVIGGHLVRDNDHPDHSRELS  
KILAKRLKQONEDLRMAESLRIPGPDQGPALGRVVRVLEQLVKAKQIENYKK



QTRNLGKHDIILRRRIENGAKELWFFLQSEKKLKNLEGNELQORHADEFLLDLGHE  
RSIMDLVYLSDGAGDWREKADLTTELVRRIITVLPKDCSKAKLVNINKGC  
GYCOLHVVYCFMAYGCTORTLILESONWRYATGWETVFRPVSETCTDPSGLSTGH  
WSGEVKDNVOVELPVIDSLHPRPPYLPVLAPEDLADRLVRHGDPAVWVSOFVKY  
LIRPQMLEKEIEBATEKLGKFKHPVIGVHVRRTDKVGEAPAFPIEETMVHVEHPQL  
LARMQVKRKYLATDDPSLLKBAKTKPNYEFISONSWSAGLHNRHYTENSRLGV  
ILDHFLSQADFLVCTTSSQVCRVAYEIMOTLHPDASANFHSLLDDIYFQGNHNOI  
AIYAHQPRADDEIPEMPGDIIGVAGNHWDGYSKGVNRKLGRTGLYPSVKVREKJETVK  
YPTYPEAEK"

## variation

2218  
/gene="FUT8"  
/note="II"  
/frequency=".5"  
/replace="c"  
2758..2763  
/gene="FUT8"  
/note="putative"  
2770..2775  
/gene="FUT8"  
/note="putative"  
2869..2874  
/gene="FUT8"  
/note="putative"

## BASE COUNT

881 a 589 c 641 g 787 t

Query Match 86.9%; Score 1501.4; DB 9; Length 2898;

Best Local Similarity 91.8%; Pred. No. 0;

Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 1 ATGCGGGCATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60  
Db 429 ATGCGGCATGAGCTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 488  
Qy 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCGAGATAATGACCCCTGATCATCTCCAGC 120  
Db 489 TTGCTGTTTATATAGTGGTCACTTGGTACGAGATAATGACCATCTCTGATCACTCTAGC 548  
Qy 121 AGAGAACTCTCCAGATTTCTGCAAGCTTGAACGCTTAAACACGAAATGAAGACTTG 180  
Db 549 CGAGAACTGTCCAAGATTTGCGAAGCTTGAACGCTTAAACACGAAATGAAGACTTG 608  
Qy 181 AGGCGAATGGCTGAGTCTCTCGAATACAGAGGCCCATTTGACCGGGGACAGCTACA 240  
Db 609 AGGCGAATGGCGAATCTCTCGGATACAGAGGCCCATTTGATCAGGGGCCAGCTATA 668  
Qy 241 GGAAGAGTCCGTGTTTATAGAAAGACAGCTTTGTTAAGGCCAAAGAACAGATGAAATPAC 300  
Db 669 GGAAGAGTACGCGTTTATAGAGAGCAGCTTTGTTAAGGCCAAAGAACAGATGAAATPAC 728  
Qy 301 AAGAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAATCTTAAGAGGAGGATTCGA 360  
Db 729 AAGAAACAAGCCAGAAATGGTCTGGGGAAGGATCATGAATCTTAAGAGGAGGATTCGA 788  
Qy 361 AATGAGCTTAAAGAGCTCTGGTGTCTTCTAACAAGCGAACTGAAGAAATTAAGCATTTA 420  
Db 789 AATGAGCTTAAAGAGCTCTGGTGTCTTCTAACAAGCGAAATTAAGAAATTAAGAACTTA 848  
Qy 421 GAAGAAATGAATCTCAAGACATGACAGATGAAATTTCTTTGGATTTAGGACACCATGAA 480  
Db 849 GAAGAAATGAATCTCAAGACATGACAGATGAAATTTCTTTGGATTTAGGACATCATGAA 908  
Qy 481 AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGAGGAGGAGGATTTGGCGT 540  
Db 909 AGGTCTATATGACGGATCTATCTACCTCAGTCAACAGATGAGGAGGAGGATTTGGCGG 968  
Qy 541 GAAAGAGGCGCAAGAGCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAAAT 600  
Db 969 GAAAGAGGCGCAAGAGCTGACAGAGCTGGTTCAGCGGAGAAATAACATATCTTCAGAAAT 1028  
Qy 601 CCTAAGGACTGCAGCAAGCCAGGAGGCTGGTGTGTAACATCAATPAAAGGCTGTGGCTTAT 660  
Db 1029 CCCAAGGACTGCAGCAAGCCAAAGAGCTGGTGTGTAATATCAACAAAGGCTGTGGCTTAT 1088

Qy 661 GGTGTCAACTCCAAACACGTGGTCTACTGTCTTTCATGATTTGCTTATGGCACCCAGCGAACA 720  
Db 1089 GGCTGTGAGCTCCAAACATGTGGTCTACTGTCTTTCATGATTTGCAATGGCACCCAGCGAACA 1148  
Qy 721 CTCACTTTGGAATCTCAGAAATTTGGCGGTATGCTACTCGTGGATGGGAGACTGTGTTTAGA 780  
Db 1149 CTCACTTTGGAATCTCAGAAATTTGGCGGTATGCTACTCGTGGATGGGAGACTGTGTTTAGA 1208  
Qy 781 CCTGTAAGTGAAGATCTACAGACAGATCTGGCCTCTCCATGGACACTGGTCAAGGTGAA 840  
Db 1209 CCTGTAAGTGAAGATCTACAGACAGATCTGGCCTCTCCATGGACACTGGTCAAGGTGAA 1268  
Qy 841 GTAATGACAAAAACATTTCAAGTGGTGGCTCCCAATTTGTAGACAGCCTCCATCTCGG 900  
Db 1269 GTGAAGGACAAAAATGTTCAAGTGGTGGCTCCCAATTTGTAGACAGCTTTCATCCCGT 1328  
Qy 901 CCTCTTACTTACCACTGGGTGTTCCAGAGAGACTTGGAGACCGACTCTTAAGAGTCCAT 960  
Db 1329 CCTCATATTTTACCCCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCAT 1388  
Qy 961 GGTGACCCCTGAGTGGTGGGTGCTCCAGTCTTGTCAAATACTTGAATTCGTCCCAACCT 1020  
Db 1389 GGTGACCCCTGAGTGGTGGGTGCTCAGTCTTGTCAAATACTTGAATTCGTCCCAACCT 1448  
Qy 1021 TGGCTGGAAGAGAAATAGAAAGCCCAAGAAAGCTTGGCTTCAAACATCCAGTTATT 1080  
Db 1449 TGGCTGGAAGAGAAATAGAAAGCCCAAGAAAGCTTGGCTTCAAACATCCAGTTATT 1508  
Qy 1081 GGAGTCCATGTACAGACGACAGAAAGTGGGAAACAGAGAGCGCTTCCACCCCATCGAG 1140  
Db 1509 GGAGTCCATGTACAGACGACAGAAAGTGGGAAACAGAGAGCGCTTCCATCCCATTTGAA 1568  
Qy 1141 GAGTACATGTTACACCTTTGAAGAACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1200  
Db 1569 GAGTACATGTTGTCATTTGAAGAACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1628  
Qy 1201 AAAAAAGAGTATATTTGGCTACTGATGATCTCTACTTTGTTTAAAGAGGAGGCAAGCAAG 1260  
Db 1629 AAAAAAGAGTATTTTGGCCACAGATGACCCCTTCTTTTAAAGAGGAGGCAAGCAAG 1688  
Qy 1261 TACTCCAATATGAATTTATTTAGTGAATCTATTTCTTTGGTCACTGAGTGGACTACAAAT 1320  
Db 1689 TACCCCAATATGAATTTATTTAGTGAATCTATTTCTTTCTGCTGAGTGGACTGCACAAAT 1748  
Qy 1321 CGGTACACAGAAATTTCACTTCGGGGTGTGATCTGAGATATACACTTCTCTCAGAGCT 1380  
Db 1749 CGATACACAGAAATTTCACTTCGGGGTGTGATCTGAGATATACACTTCTCTCAGAGCA 1808  
Qy 1381 GACTTTCTAGTGTCTACTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCAA 1440  
Db 1809 GACTTCTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCAA 1868  
Qy 1441 ACCCTGCTATCTGATGCTCTCGAACTTCCATTTCTTTGGATGATCATCTACTATTTTGG 1500  
Db 1869 ACCTTACATCTGATGCTCTCGAACTTCCATTTCTTTAGATGATCATCTACTATTTTGG 1928  
Qy 1501 GGCCAAATGGCCACATCAGATTTGCTTTTATCTCACAACCTCGAACTGGAAGAGAA 1560  
Db 1929 GGCCAGAAATGGCCACATCAAAATTTGCCATTTATGCTCACAACCCCGAACTGCAGATGAA 1988  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCATTTGGATGGTTATTCT 1620  
Db 1989 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAATCATTTGGATGGCTATTCT 2048  
Qy 1621 AAAGGTATCAACAGAAATTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 2049 AAAGGTATCAACAGAAATTTGGGAAAGGAGCGGCTTATATCCCTCTCAAAAGTTCGAGAG 2108  
Qy 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAAGCTGAAAAATA 1727  
Db 2109 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAGGCTGAGAAATA 2155

RESULT 11	HSAS39536	2902 bp	mRNA	linear	PRI 03-FEB-2003	61	TTGTTATTATATAGTGGTCAATTTGGTTCGAGATAATGACACACCTCTGATCACTCCAGC	120	
LOCUS	HSAS39536					386	TTGCTGTTTATATAGTGGTCACTTGGTACGAGATAATGACCATCTGATCACTCTGAC	445	
DEFINITION	HSAS39536	2902 bp	mRNA	linear	PRI 03-FEB-2003	121	AGAGAACTCTCAAGATTCTTGGCAAGCTTGAACCGCTTAAAAACAGCAAAATGAAGACTTG	180	
ACCESSION	AJ539536					446	CGAGAACTGTCCAGATTCTGGCAAGCTTGAACGCTTAAAAACAGCAGATGAAGACTTG	505	
VERSION	AJ539536.1	GI:28200809				181	AGCGCAATGGTGTAGTCTCTCCGAATACCAAGAGGCCCATTTGACACAGGGGACAGCTACA	240	
KEYWORDS	alpha6-fucosyltransferase; FUT8 gene.					506	AGCGCAATGGCGGAATCTCTCCGATACCAAGAGGCCCTTATTTGATCAGGGGCGCAGCTATA	565	
SOURCE	Homo sapiens					241	GGAGAGTCCGTGTTTATAGAGACAGCTGTTTAAGGCCCAAGAACACAGATTGGAATAATTAC	300	
ORGANISM	Homo sapiens					566	GGAGAGTACCGCTTTTGAAGAGCAGCTGTTTAAGGCCCAAGAACACAGATTGGAATAATTAC	625	
REFERENCE	1	Mollicone, R., Michalski, J.C., Bauvy, C., Cailletau-Thomas, A., Candelier, J.J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.				301	AAGAAACAAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTTAAAGAGAGGATTGAA	360	
AUTHORS						626	AAGAAACAAGCCAGCAAGAAATGGTCTCGGGAAGGATCATGAAATCTCTGAGGAGGAGATTGAA	685	
TITLE	Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis					361	AATGAGCTTAAAGAGCTCTGTTTTTCTACAAAGCGAACTGAAGAAATTTAAAGCATTTTA	420	
JOURNAL	Unpublished					686	AATGAGCTTAAAGAGCTCTGTTTTTCTACAGAGTGAATTTGAAGAAATTTAAAGAACTTA	745	
REFERENCE	2 (bases 1 to 2902)					421	GAAGCAATGAACCTCCAAAGACATGCAGATGAATCTTTTGGATTTTAGGACACCATGAA	480	
AUTHORS	Oriol, R.					746	GAAGCAATGAACCTCCAAAGACATGCAGATGAATCTTTTGGATTTTAGGACATCATGAA	805	
TITLE	Submitted (31-JAN-2003) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE					481	AGGTCTATCATGACAGATCTATACTACTCAGTCAAAACAGATGGAGCAGGGGATTTGGCGT	540	
JOURNAL	Vaillant-Couturier, 94807, FRANCE					806	AGGTCTATTAATGACGGATCTATACTACTCAGTCAAGACATGGAGCAGGTGATTGGCGG	865	
FEATURES	Location/Qualifiers					541	GAAGAGGCGCAAGATCTGACAGAGCTGGTTCAGCGGAGATAAACAATATCTCCAGAAT	600	
source	1. .2902					866	GAAGAGGCGCAAGATCTGACAGAACTGGTTTCAGCGGAGATAAACAATATCTCCAGAAT	925	
gene	/organism="Homo sapiens"					601	CCTAAGGACTGCAGCAAGCCAGGAAGCTGGTGTGTAACATCAATAAGGCTGTGGCTAT	660	
CDS	/mol_type="mRNA"					926	CCCAAGGACTGCAGCAAGCCAGGAAGCTGGTGTGTAATCAACAAGGCTGTGGCTAT	985	
	/db_xref="taxon:9606"					661	GGTTCTCAACTCCATCAGTGGTCTACTGTTTTCATGATGCTTATGGCACCACCGAGAAC	720	
	/chromosome="14"					986	GGTCTCAGCTCCATCATGTGCTCTACTGCTTCATGATTCATATGSCACCCAGCGAAC	1045	
	/map="14q23.2"					721	CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGCTGGTGGAGACTGTATTTAG	780	
	/clone="CD-L2"					1046	CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGCTGGTGGAGACTGTATTTAG	1105	
	/tissue type="whole embryo"					781	CCTGTAAGTGAAGATGTACAGACAGATCTGGCGCTCTCCACTGAGCACTGGTCAAGTGAA	840	
	/dev stage="40-70 days"					1106	CCTGTAAGTGAAGATGTACAGACAGATCTGGCGCTCTCCACTGAGCACTGGTCAAGTGAA	1165	
	/country="France:Paris"					841	GTAAATGACAAAAACATTTCAAGTGGTTCGAGCTCCCGCTTCCTTAGACAGCTCCATCTCTCG	900	
	/EC number="2.4.1.68"					1166	GTGAAGGACAAAATGTTCAAGTGGTTCGAGCTTCCTCCACTGAGCACTGGTCAAGTGCAAT	1225	
	/function="adds fucose in alpha1,6 linkage to the internal GlcNAc of the core chitobiose"					901	CCTCTTACTTACCACTGGCTGTTTCCAGAGACCTTCGACACCGACTCTTCAAGAGTCCAT	960	
	/codon_start=1					1226	CCTCTCATATTTACCCCTTGGCTGTATCCAGAGACCTTCGACATCGACTTGTACGAGTGCAAT	1285	
	/evidence=experimental					961	GGTGACCTTCAGTGTGGTGGTGTCCAGTTCCTCAAACTACTTGAATTCGTCCACACACCT	1020	
	/product="alpha6-fucosyltransferase"					1286	GGTGACCTTCAGTGTGGTGGTGTCTCAGTTCCTCAAACTACTTGAATTCGTCCACACACCT	1345	
	/protein_id="CAD62445.1"					1021	TGGCTGGAAAAAGAAATAGAAAGACCAAGAGCTTGGCTTTCAAACTATCCAGTTTATT	1080	
	/db_xref="GI:28200810"					1346	TGGCTTAGAAAAAGAAATAGAAAGACCAAGAGCTTGGCTTTCAAACTATCCAGTTTATT	1405	
	/translation="MRPWTGSRWIMLILFAWGTLLFVIGGLVLRNDHPDHSRELS KILAKLERLKQNEDLRRMSSLRIPEGPIDQGAIRVRVLEQLVKAKEQIENYKK QRTNGKGDKEILRRRIENGAKELFFLOSELKKNLKNELGRHDEFLDLGHHE RSIMTDLYSLDGDAGWREKAKDLTELVRITLYLQNPDKCSKAKLVLVCKNGC GYCQLHVYQFMIAVTGRTLLILESONWRYATGWTVPVPSVETCTDRSGISTGH WSGEVKDKNVQVVELPIVDLSLHPRPPYLAPEVDLADRLVRHGDPAVMWVSQFVKY LIRPQWPKETEEATKLGFKHPVIGVHVVRTDKVGTAAAFHPIDBYMVHVEHFL LARMQVKKRYLATDPSLLKEATKYPNYEPISDNSISWAGLHNRVTENSLRGV ILDHPLSDAFELVTFSSQVRVAIEIMQTLHPDASANFSLDDIIFYGGQNAHQI A1YAHQPRPADSIPMEPGDIIIGVAGNHWGDSKGVNRKLGRTGLYPSKYREKIETVK YTPPEAEK"					1081	GGAGTCCATGTTCAGACGCGACAGACAAAGTGGGAACAGAACGAGCTTCCACCCCATCGAG	1140	
	/gene="FUT8"					1406	GGAGTCCATGTTCAGACGCGACAGACAAAGTGGGAACAGAACGAGCTTCCATCCCATTTGAA	1465	
	/db_xref="taxon:9606"								
	/map="14q23.2"								
	/clone="CD-L2"								
	/tissue type="whole embryo"								
	/dev stage="40-70 days"								
	/country="France:Paris"								
	/EC number="2.4.1.68"								



Qy	301	AAGAAA	CAAGCT	AGAAAT	TGGTCT	GGGGAAG	GATCAT	GAAATCT	TTAAAG	AGGAGAT	TGAA	360
Db	729	AAGAAA	CAGAC	CAGAAA	TGGTCT	GGGGAAG	GATCAT	GAAATCT	TTAAAG	AGGAGAT	TGAA	788
Qy	361	AATGGAG	CTAAGAG	CTCTGG	TTTTTCT	CAAAAG	CGAAT	CTGAAAG	AAAT	TAAGCA	TTTA	420
Db	789	AATGGAG	CTAAGAG	CTCTGG	TTTTTCT	CAAAAG	CGAAT	CTGAAAG	AAAT	TAAGCA	TTTA	848
Qy	421	GAAGGAAA	TGAAT	CTC	CAAGAC	TGATG	CAAAAT	TTCTTT	TGGAT	TTAGGAC	CAATGAA	480
Db	849	GAAGGAAA	TGAAT	CTC	CAAGAC	TGATG	CAAAAT	TTCTTT	TGGAT	TTAGGAC	CAATGAA	908
Qy	481	AGGTC	TATCAT	GA	CAGATCT	ATACT	CACT	CAGTCA	AAACAG	ATGAG	CAGGGGAT	540
Db	909	AGGTC	TATA	TG	CGGATCT	ATACT	CACT	CAGTCA	GAAG	ATGAG	CAGGGGAT	968
Qy	541	GAAGGAGG	CCAAAG	ATCT	CA	CAGAGCT	TGGT	CCAGCG	GAGAA	TAA	CATATCTCC	600
Db	969	GAAGGAGG	CCAAAG	ATCT	CA	CAGAGCT	TGGT	CCAGCG	GAGAA	TAA	CATATCTCC	1028
Qy	601	CCTAAG	GACTC	CAG	CAAA	GGCAG	GAAGCT	TGGT	GTAA	CACTCA	ATAAAGGCTGTGG	660
Db	1029	CCCAG	AGACT	CG	CAAA	AGCCAA	AGAGCT	TGGT	GTAA	CACTCA	ATAAAGGCTGTGG	1088
Qy	661	GGTTGT	CAACT	CCAT	CA	CGTGGT	CTACT	TGTTT	CAATG	ATGCTT	ATGCA	720
Db	1089	GGCTGT	CAGCT	CCAT	CA	TGTTGGT	CTACT	TGTTT	CAATG	ATGCTT	ATGCA	1148
Qy	721	CTCAT	CTTTG	GAAT	CT	CAGAA	TTGG	CGCTAT	GC	TACT	GGTGGATGGGAG	780
Db	1149	CTCAT	CTTTG	GAAT	CT	CAGAA	TTGG	CGCTAT	GC	TACT	GGTGGATGGGAG	1208
Qy	781	CCTGT	TAAGT	GAG	ACAT	GT	PAC	AGACAG	ATCT	GG	CGCTCT	840
Db	1209	CCTGT	TAAGT	GAG	ACAT	GT	PAC	AGACAG	ATCT	GG	CGCTCT	1268
Qy	841	GTAAT	GACAAAA	CA	TT	CAAGT	GGT	CGAGCT	CC	CCCAT	TTGTAG	900
Db	1269	GTGAAG	GACAAAA	CA	TT	CAAGT	GGT	CGAGCT	CC	CCCAT	TTGTAG	1328
Qy	901	CCTCCT	TACT	TAC	CA	CTGGCT	GT	TC	AGAG	ACCT	TTGAG	960
Db	1329	CCTCCT	TACT	TAC	CA	CTGGCT	GT	TC	AGAG	ACCT	TTGAG	1388
Qy	961	GGTCA	CCCT	CG	AGT	GTGG	GT	CTCC	AGT	TTG	CAAACT	1020
Db	1389	GGTCA	CCCT	CG	AGT	GTGG	GT	CTCC	AGT	TTG	CAAACT	1448
Qy	1021	TGGCT	TG	AAAA	AGG	AAAT	TAG	AA	AG	AGCC	CAAG	1080
Db	1449	TGGCT	TG	AAAA	AGG	AAAT	TAG	AA	AG	AGCC	CAAG	1508
Qy	1081	GGAGT	CCAT	GT	CA	GC	CA	CAG	CA	AA	AGTGG	1140
Db	1509	GGAGT	CCAT	GT	CA	GC	CA	CAG	CA	AA	AGTGG	1568
Qy	1141	GAGT	TAC	AT	GGT	TG	AG	AC	AT	TTT	CAG	1200
Db	1569	GAGT	TAC	AT	GGT	TG	AG	AC	AT	TTT	CAG	1628
Qy	1201	AAAAA	AG	AGT	AT	AT	CT	GG	CT	AC	TG	1260
Db	1629	AAAAA	AG	AGT	AT	AT	CT	GG	CT	AC	TG	1688
Qy	1261	TACT	CC	AA	TT	AT	TG	AT	GA	TAA	CT	1320
Db	1689	TACCC	AA	TT	AT	TG	AT	GA	TAA	CT	CT	1748
Qy	1321	CGGT	AC	AC	AG	AA	AT	T	CA	CT	CG	1380
Db	1749	CGAT	AC	AC	AG	AA	AT	T	CA	CT	CG	1808
Qy	1381	GACT	TT	CT	AG	T	GT	GA	CT	TT	TT	1440

[illegible]

```
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein id="CAA76987.1"
/db_xref="GI:3451267"
/translaton="MRPTGSRWIMLILFWGCTLLFYIGHLVRDNDHPHSRELS
KILAKLERKQONEDLRMAESLRIPEGPIDOGPAIGRVRVLESQLVKAEQIENYKK
RTNGLGKHDEILRRIRFESSGAKELWFFLOSELKKNLEGNELQRHADFLLDLGHHE
RSITNDLYLSQTDGAGDREKEAKDLTELVRITVILQNPDKCSKAKLVCNINKGC
GYGQHLHVYCFMIAVCTQRTLLILESONWRYATGWEVFPVSECTDRSGISTGH
WSGEVKNQVOVELPIVDSLHPRPVPLPLAVPEDLADLRVVRHGDPAVMVSOFVKY
LIRPQWLEKIEBEATKGLGFKHPVIGVHVVRTDKVTEAAPHIEEYMHVHEHFQL
LARMQVKKRVIYLAIDPDSLLKEAKTKIPNTEFISONSISWAGLNRHYTENSLRGV
LIDHLFSQADPLVCTGSSQVRVAYEIMQTLHPDASNFHSLDDIYFPGQNAHQI
AIYAHOPRTADEIPMEPGDIIICVAGNHWGYSKGVNRKLGRTGLYPSYKVKREKLETKV
YPTVPEAEK"
variation 2506
/gene="FUT8"
/note="II"
/frequency=".5"
/replace="c"
3046..3051
/gene="FUT8"
/polyA_signal
/note="putative"
3058..3063
/gene="FUT8"
/note="putative"
3157..3162
/gene="FUT8"
/note="putative"
BASE COUNT 910 a 710 c 721 g 840 t 5 others
ORIGIN
Query Match 86.9%; Score 1501.4; DB 9; Length 3186;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1 ATCGGGGATGACGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 60
DB 717 ATCGGCCATGACGTGGTTCCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGACC 776
QY 61 TTGTTATTTTATATAGTGGTTCATTTGGTTGAGATAATGACCCCTGATCATCTCCAGC 120
DB 777 TTGCTGTTTTATATAGTGGTTCATTTGGTACGAGATAATGACCATCTCTGATCACTCTAGC 836
QY 121 AGAGAACTCTCAAGATCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTG 180
DB 837 CGAGAACTGTCCAAGATCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTGCAAGCTTG 896
QY 181 AGGCGAATGGCTGAGTCTCTCGAATACAGAGAGGCCCATTTGACCGGGGACAGCTACA 240
DB 897 AGGCGAATGGCGAATCTCTCGGATACAGAGAGGCCCATTTGATCAGGGGCCAGCTATA 956
QY 241 GGAAGAGTCCGTGTTTGAAGAACAGCTTGTTAAGGCCCAAAGAACAGATGAAATTTAC 300
DB 957 GGAAGAGTACGCGTTTGAAGAGCAGCTTGTTAAGGCCCAAAGAACAGATGAAATTTAC 1016
QY 301 AAGAAACAGCTAGAAATGGTCTGGGAGAGATCATGAATCTTAAGAGAGAGATTTGAA 360
DB 1017 AAGAAACAGACCAAGAAATGGTCTGGGAGAGATCATGAATCTCTGAGGAGAGATTTGAA 1076
QY 361 AATGAGCTTAAGAGCTCTGGTTTTTCTTCAAGAGCAACTGAAGAAATTAAGCATTTA 420
DB 1077 AATGAGCTTAAGAGCTCTGGTTTTTCTTCAAGAGCTGAATTTGAAGAAATTAAGNACTTA 1136
QY 421 GAAGAAATGAATCTCAAGAACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 480
DB 1137 GAAGAAATGAATCTCAAGAACATGCAGATGAAATCTTTTGGATTTAGGACATCATGAA 1196
QY 481 AGGTCTATCATGACAGATCTATATCTTACCTCAGTCAAAACAGATGAGAGAGGGATTTGGCGT 540
DB 1197 AGGTCTATATATGACGGATCTATATCTTACCTCAGTCAAGAGATGAGAGAGGGATTTGGCGG 1256
QY 541 GAAAGAGGCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 600
```

```
Db 1257 GAAAAGAGGCCAAGAGATCTGACAGAAATCGTTCTCAGCGGAGAAATAATATCTTCAGAAAT 1316
QY 601 CCTAAGGACTGACACAAGCCAGGAAGCTGGTGTGAACATCAATAAAGGCTGTGGGTAT 660
DB 1317 CCAAGGACTGACACAAGCCAAAAGCTGGTGTGAATAATCAACAAAAGGCTGTGGGTAT 1376
QY 661 GGTGTCAACTCCATCAGCTGGTGTCTACTGTTCATGATTTGCTTATGACACCCAGCAACA 720
DB 1377 GGCTGTGAGCTCCATCATGTGGTCTACTGTTCATGATTTGCTTATGACACCCAGCAACA 1436
QY 721 CTGATCTTGAATCTCAGAAATTTGGCGCTAGTCTACTGTGTGATGGAGACTGTGTTTGA 780
DB 1437 CTGATCTTGAATCTCAGAAATTTGGCGCTAGTCTACTGTGTGATGGAGACTGTATTTAGG 1496
QY 781 CCTGTAAGTGAGACATCTACAGACAGATCTGGCGCTCTCCACTGACACTGTGTGAGTCAA 840
DB 1497 CCTGTAAGTGAGACATCTACAGACAGATCTGGCGCTCTCCACTGACACTGTGTGAGTCAA 1556
QY 841 GTAAATGACAAAAACAATTCAAGTGGTGGAGCTCCCATTTGTAGACAGCTCCATCTCTCGG 900
DB 1557 GTGAAGGACAAAAATGTTCAAGTGGTGGAGCTCCCATTTGTAGACAGTCTTCATCTCCCGT 1616
QY 901 CCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGGAGACCGACTCTTAAGAGTCCAT 960
DB 1617 CCTCCATATTTACCCCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTTGTACGAGTGCAT 1676
QY 961 GGTGACCTCGAGTGTGGTGGTCTCCAGTTTCTCAATACTTGTGTCGTCACAACT 1020
DB 1677 GGTGACCTCGAGTGTGGTGGTCTCTCAGTTTGTCAATACTTGTGATCCGCCCCACAGCT 1736
QY 1021 TGGCTGAAAAGAAAATAGAAAGAACCAAGAGCTTGGCTTCAAAATCCAGTATTT 1080
DB 1737 TGGCTAGAAAAGAAAATAGAAAGAACCAAGAGCTTGGCTTCAAAATCCAGTATTT 1796
QY 1081 GGAGTCCATGTACAGACACAGACAAAAGTGGGAAACAGAGAGCTTCCACCCCATCGAG 1140
DB 1797 GGAGTCCATGTACAGACACAGACAAAAGTGGGAAACAGAGAGCTTCCATCCCATTTGAA 1856
QY 1141 GAGTACATGTACAGCTTGAAGACATTTTCAGCTTCTCGCAGCAGATCGAATCGAT 1200
DB 1857 GAGTACATGTGTCATGTGAAGAACATTTTCAGCTTCTTCGACGAGATGCAAGTGGAC 1916
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTCTACTTTGTTAAAGAGGCAAGAACAAAG 1260
DB 1917 AAAAAAGAGTATTTTGGCCACAGATGACCCCTTCTTTTAAAGAGGCAAGAACAAAG 1976
QY 1261 TACTCCAAATTATGAATTTATTTAGTGATACTTATTTCTTTGGTCAAGCTGGAATACAAAT 1320
DB 1977 TACCCCAATTATGAATTTATTTAGTGATACTTATTTCTTCTGCTGAGCTGGACTGCACAT 2036
QY 1321 CGGTACACAGAAATTTCACTTCGGGGTGTGATCTCGATATACATTTCTCTCACAGCT 1380
DB 2037 CGATACACAGAAATTTCACTTCGGGAGTGTATCTCGATATACATTTTCTCTCAGGCA 2096
QY 1381 GACTTTCTAGTGTCTACTTTTTTCACTCCAGCTCTGTGGTGTGTTATGAATCATGCAA 1440
DB 2097 GACTTTCTAGTGTGTACTTTTTTCACTCCAGCTCTGTGGTGTGTTATGAATTTATGCAA 2156
QY 1441 ACCCTGCTATCTGATGCTCTCGAACTTCATTTCTTTGGATGATCATCTACTATTTTGA 1500
DB 2157 ACCTACATCTGATGCTCTCGAACTTCATTTCTTTTAGATGATCATCTACTATTTTGG 2216
QY 1501 GGCCAAATGCCACATCAGATTTGCTGTTTATCTCTCAAAACCTCGAATCGAAGAGAA 1560
DB 2217 GGCCAGAAATGCCACATCAAATTTGCAATTTATGCTCACCAACCCCGCAATCGCAGATGAA 2276
QY 1561 ATTTCAATGGAACTGGAGATCATTTGGTGTGGCTGAAACCATTTGGATGGTATTTCT 1620
DB 2277 ATTTCCATGGAACTGGAGATCATTTGGTGTGGCTGAAATCATTTGGATGGCTATTTCT 2336
QY 1621 AAAGGTATCAACAGAAACTTTGGAAACACAGCTTATATCTCTCTCAAAAGTCCGAGAG 1680
```

```
Db 2337 AAGGTGTCAACAGGAATTTGGGAAGGAGGCGCTATATCCCTCTCAAAAGTTCCAGAG 2396
Qy 1681 AAGTAGAAACAGTCAAGTATCCCATATCTCTGAAGTGAATAAATA 1727
Db 2397 AAGTAGAAACAGTCAAGTATCCCATATCTCTGAAGTGAATAAATA 2443

RESULT 14
LOCUS HSY17979 3280 bp mRNA linear PRI 06-NOV-1998
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase
transcript A2.
ACCESSION Y17979
VERSION Y17979.1 GI:3451268
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Cailleau,A., Balanzino,L., Candelier,J.J., Oriol,R. and
Mollicone,R.
TITLE Differential splice variants of human FUT8 embryonic cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3280)
AUTHORS Cailleau,A.L.V.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailleau, INSERM U178, 16 av P.V.
Couturier, 94807 Villejuif Cedex, FRANCE
COMMENT Location/Qualifiers
FEATURES
source
1..3280
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23"
/clone="FUT8 transcript A2"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
1..3280
/gene="FUT8"
1..389
/gene="FUT8"
/note="A"
390..488
/gene="FUT8"
/note="C"
489..>2444
/gene="FUT8"
/note="D"
717..2444
/gene="FUT8"
/EC_number="2.4.1.68"
/note="allele 1"
/codon_start=1
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CAA76988.1"
/db_xref="GI:3451269"
/db_xref="SPTREMBL:O00235"
/translation="MRPWTGSRWIMLILFAWGTLLFYIGHLVRDNDHPDHSRELS
KILAKRLKQONEDLRMAESLRIPGPIQGPAGIRVRVLEQLVKAQIENYKK
QTRNLGKHDETLRRIRINGAKELWFFLQSELKKLNLEGNELORHAEFLDLGHHE
RSIMTDLVLSQTDGADWREKADLTJLVQRITLYLQNPKDCSKAKLVNINKGK
GYGQHLHVVICFMATYGTQRTLLILESONRWYATGNETVFRPSETCTDSRGISTGH
WSGEVKDNQVVELPIVDSLHPRPPIPLAVPEDLRLVRVHGDPVAVWVSQFVKY
LIRPQWLKEETEELKKLGKHPVIGVVRRTDKVTEAAPHPIEYMYVENSFLQ
LARRMQVKRVLATDPSLLKEAKTPYNEFTSDNSISWSAGLHNRVYENSIRGV
ILDHFLQADFLVCTFSSQVCRVAYEIMQTLHPDASANFSLDDIYYFGQNAHNOI
ALYAHQPTADEIPMEPGDILIGVAGNHDGYSKGVNKLGRITGLYPSYKVKREKIEYK
YPTYPEAK"
2506
/variation
/gene="FUT8"
```

```
/note="II"
/frequency=".5"
/replace="c"
3046..3051
/gene="FUT8"
/note="putative"
3058..3063
/gene="FUT8"
/note="putative"
3157..3162
/gene="FUT8"
/note="putative"
BASE COUNT 934 a 718 c 743 g 880 t 5 others
ORIGIN
Query Match 86.9%; Score 1501.4; DB 9; Length 3280;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy 1 ATGCGGCATGGAAGTCTCTGCGGTGGATTATGCTCATCTTTTGCCTGGGGACC 60
Db 717 ATGCGGCATGGAAGTCTCTGCGGTGGATTATGCTCATCTTTTGCCTGGGGACC 776
Qy 61 TTGTTATTTATATAGTGGTCAATTCGTTTCGAGATATGACCCCTGATCACTCCAGC 120
Db 777 TTGCTGTTTTATATAGTGGTCACTTGTGACGATATGACCATCTCTGATCACTTAGC 836
Qy 121 AGAGAACTCTCCAAAGATTCTTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180
Db 837 CGAGAACTGTCCAAAGATTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 896
Qy 181 AGCGAATGGTGTAGTCTCTCCGAATACCGAAGGCCCCATTTGACAGGGGACAGCTACA 240
Db 897 AGCGAATGGCGGAATCTCTCCGATACCGAAGGCCCCATTTGATCAGGGGCCAGCTATA 956
Qy 241 GGAAGAGTCCGTGTTTGTAGAGAACAGCTTCTTAAGGCAAAAGAACAGATGAAATTTAC 300
Db 957 GGAAGAGTACCGCTTTTGTAGAGAGCAGCTTCTTAAGGCAAAAGAACAGATGAAATTTAC 1016
Qy 301 AAGAAAACAAGCTAGAAATGGTCTGGGAAGATCATGAAATCTTTAAGAAAGAGGATTCGAA 360
Db 1017 AAGAAAACAGCAGAAATGGTCTGGGAAGATCATGAAATCTCTGAGGAGGAGATTGAA 1076
Qy 361 AATCGAGCTAAGAGCTCTGGTTTTTCTCAAGCGAAGTGAAGAAATTTAAGCATTTA 420
Db 1077 AATGAGCTAAGAGCTCTGGTTTTTCTCAAGAGTGAATTTGAAGAAATTTAAGAACTTA 1136
Qy 421 GAAGGAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTTAGGACACCATGAA 480
Db 1137 GAAGGAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTTAGGACATCATGAA 1196
Qy 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGGAGCAGGGGATTTGGCGT 540
Db 1197 AGGTCTATATGACGGATCTATACCTCAGTCAAGACAGATGGAGCAGGTGATTGGCGG 1256
Qy 541 GAAAAGAGGCCAAAGATCTGACAGCTGTGTGTCAGCGGAGAAATTAACATATCTCCAGAT 600
Db 1257 GAAAAGAGGCCAAAGATCTGACAGAACTGTGTTTCAGCGGAGAAATTAACATATCTCTCAGAAT 1316
Qy 601 CCTAAGGACTGCACAAAGCCAGGAAGCTGGTGTACATCAATAAAGGCTGTGGCTAT 660
Db 1317 CCCAAGGACTGCACAAAGCCAAAGCTGGTGTATATATCAAAAGGCTGTGGCTAT 1376
Qy 661 GGTGTCAATCCATCAGCTGGTCTACTGTTTTCATGATTGCTTATGGCACCCAGCGAACA 720
Db 1377 GGCTGTAGCTCCATCATGTGCTCTACTGCTTCATGATTGATATGGCACCCAGCGAACA 1436
Qy 721 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTGTGATGGAGACTGTGTTTGA 780
Db 1437 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTGTGATGGAGACTGTATTTAGG 1496
Qy 781 CCTGTAGTGAACATGTACAGACAGATCTCGCCTCTCCACTGGACACTGGTCAAGTGCA 840
```

Db	1497	CCTGTAAGTGAGACATGCACAGACAGATCTGGCAATCTCCACTGGACACTGGTGCAGTGAA	1556	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Qy	841	GTAAATGACAAAAACATTCAGTGTGCGAGCTCCCAATGTGTAGACAGAGCTCCATCCTCGG	900	1 Mollicone, R., Michalski, J. C., Bauvy, C., Caillean-Thomas, A., Candelier, J. J., Martinez-Duncker, I., Breton, C., Codogno, P. and Oriol, R.
Db	1557	GTGAAGGACAAAATGTTCAAGTGTGAGGCTTCCCATGTAGACAGTCTCATCCCGCT	1616	Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis
Qy	901	CCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCGACTCCTTAAGAGTCCAT	960	Unpublished
Db	1617	CCTCCATATTACCTTGGCTGTACAGAGACCTCGCAGATCGACTTGTACGAGTGCAT	1676	2 (bases 1 to 3568)
Qy	961	GGTAGCCCTGCAGTGTGGTGGTCCAGTTTGTCAATATCTGATTCGTCACAACT	1020	Direct Submission
Db	1677	GGTAGCCCTGCAGTGTGGTGGTGTCTCAGTTTGTCAATATCTGATTCGTCACAACT	1736	Submitted (06-JAN-2003) Oriol R., US04, INSERM, 16 Avenue Paul Vallant-Couturier, Villejuif, 94807, FRANCE
Qy	1021	TGGCTGGAAGAAATAGAGAAGCCACCAAGAGCTTGGCTTCAAAATCTCCAGTTATT	1080	related splice variants AJ536053.1, AJ536054.1 and AJ536056.1.
Db	1737	TGGCTAGAAAAGAAATAGAGAAGCCACCAAGAGCTTGGCTTCAAAATCTCCAGTTATT	1796	Location/Qualifiers
Qy	1081	GGAGTCCATGTGACGACGACAAAAGTGGGAAACAGAGCAGCTTCCACCCCATCGAG	1140	1. 3568 /organism="Homo sapiens"
Db	1797	GGAGTCCATGTGACGACGACAAAAGTGGGAAACAGAGCAGCTTCCATCCCATCGAA	1856	/mol_type="mrna"
Qy	1141	GAGTACATGTTACGTTGAAGAACATTTTCAGTCTTCGACGAGCAATGCAAGTGCAT	1200	/db_xref="taxon:9606"
Db	1857	GAGTACATGTTGATGTTGAAGAACATTTTCAGTCTTCGACGAGCAATGCAAGTGCAC	1916	/chromosome="14"
Qy	1201	AAAAAAGAGTATATCTGCTACTGATGATCTCTGTTTAAAGAGGCAAGACAAAG	1260	/map="14Q23.2"
Db	1917	AAAAAAGAGTATTTGGCCACAGATGACCTTCTTTTAAAGAGGCAAGACAAAG	1976	/clone="E2D-L1"
Qy	1261	TACTCCATTTATGAATTTATTAGTGATAAATCTATTTCTGGTACGCTGGACTACAAAT	1320	/tissue_type="whole embryo"
Db	1977	TACCCCAATTATGAATTTATTAGTGATAAATCTATTTCTGGTACGCTGGACTACAAAT	2036	/dev_stage="40-70 days"
Qy	1321	CGGTACACAGAAAATTCATCTCGGGGTGTGATCTGGATATACATTTCTCTCACAGCT	1380	/country="France:Paris"
Db	2037	CGATACACAGAAAATTCATCTCGGGGTGTGATCTGGATATACATTTCTCTCACAGCA	2096	1. 3568 /gene="FUT8"
Qy	1381	GACTTCTAGTGTGACTTTTTTCATCCAGGTCGTGCGGTGCTTATGAATCATGCAA	1440	/EC_number="2.4.1.68"
Db	2097	GACTTCTAGTGTGACTTTTTTCATCCAGGTCGTGCGGTGCTTATGAATCATGCAA	2156	/functions="adds fuc in alphas1,6 to the internal GlcNAc of the core chitobiose"
Qy	1441	ACCTGTCATCTGATGCTCTCGCAACTTCCATTTCTTTGGATGACATCTACTATTTTGA	1500	/codon_start=1
Db	2157	ACACTACATCTGATGCTCTCGCAACTTCCATTTCTTTAGATGACATCTACTATTTGG	2216	/evidence=experimental
Qy	1501	GGCCAAAATGCCCAACAATCAGATTCGTGTTATCTCTCAAACTCGCAACTGAAGAGAA	1560	/product="alpha6-fucosyltransferase"
Db	2217	GGCCAGAATGCCCAACAATTCGCAATTTATGCTCACCACCCGCAACTGCAGATGAA	2276	/protein_id="CAD59926.1"
Qy	1561	ATTCCAATGGACCTGGAGATATCATTTGTTGGCTGGAAACCATTTGGATGTTATCT	1620	/db_xref="GI:27552400"
Db	2277	ATTCCAATGGACCTGGAGATATCATTTGTTGGCTGGAAATCATTTGGATGTTATCT	2336	/translation="MRPWTGSRWIMLIPAWGTLFLVIGHVLVDNDHPDHSRELS KILALRLKQONEDLERMAESLRIPSGPIDQGAIGRVRVLEQLVKAKQEIENYKK RNTNGLGRHEILRRIRENGAKELWFFLOSELKKLKNELGRLQRADEFLDLGHKE RSTMDLYLSUTDAGDWEKEADLTLEORRITLYQNPDKCSKAKGVCNINNGC GYGQLHHVYCFMAYLPVDSLHPPVPLAVPEDLADRLVRVHGDPAVMVVSQFVK WSGVBKDNQVQVSLPIVDSLHPPVPLAVPEDLADRLVRVHGDPAVMVVSQFVK LIRQPWLEKEIEATKLGKHPVIGVHVRTDKVTEAAPHPIEEVMHVESHFOL LARMQVDKRVVLATDDPSLLKEATKYPNYEFISNSISWSAGLHNRVTENSIRGV ILDIHFUSQADFLVCTSQVCRVAYEIMQTLHPDASANFHSLLDDIYFFGQNAHNQI AIYAHQPRTADEIPMEPGDIIGVAGNHWGDSYSGVNRKLGRTGLDYPSTYKVKREKLETVK YPTYPEAEK"
Qy	1621	AAAGGTATCAACAGAAAACTTGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG	1680	polyA_signal 3538..3543
Db	2337	AAAGGTATCAACAGAAAACTTGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG	2396	/gene="FUT8"
Qy	1681	AAGATAGAAAACAGTCAAGTATCCCAATATCTCTGAAGCTGAAAAATA	1727	3553
Db	2397	AAGATAGAAAACAGTCAAGTATCCCAATATCTCTGAAGCTGAAAAATA	2443	/BASE COUNT 1054 a 751 c 788 g 975 t
RESULT 15	HSAS36055	3568 bp	linear	Query Match 86.9%; Score 1501.4; DB 9; Length 3568;
LOCUS	HSAS36055			Best Local Similarity 91.8%; Pred. No. 0;
DEFINITION	Homo sapiens mRNA for alpha6-fucosyltransferase (FUT8 gene), splice variant B5.			Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
ACCESSION	AJ536055			
VERSION	AJ536055.1	GI:27552399		
KEYWORDS	alpha6-fucosyltransferase; FUT8 gene.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			

[illegible]

Db	2358	TACCCCAATTATGAAATTTATTAGTGATAA	CTATTATTTCTGGTCAGCTGCAGTCA	24117
Qy	1321	CGGTACACAGAAAATTCAC	TTCCGGGGTGTGATCCTCGATATACACTTTCTCTCACAGGCT	1380
Db	2418	CGATACACAGAAAATTCAC	TTCTGTGGAGTGATCCTCGATATACATTTTCTCTCAGGCA	24777
Qy	1381	GACTTTCTAGTGTGTACTTTTT	CATCCCAAGGTCGTGCGGTGTCTTATGAAATCATGCAA	1440
Db	2478	GACTTCTAGTGTGTACTTTTT	CATCCCAAGGTCGTGAGTGTCTTATGAAATATGCAA	2537
Qy	1441	ACCTGCATCCTGATGCCTCTG	CGAACTTCCATTTCTTGGATGACATCTACTATTTTGA	1500
Db	2538	ACACTACATCCTGATGCCCT	CTGCAAACTTCCATTTCTTAGATGACATCTACTATTTTGG	2597
Qy	1501	GGCCAAAATGCCCAACATCAG	ATGTGTTTATCTTCACAAACTCGAACTGAAGAGGAA	1560
Db	2598	GGCCAGATGCCCAACATCAA	ATTGCTCACCACCCCGAACTGCAGATGAA	2657
Qy	1561	ATTCCAATGAACTCGAGATAT	CAATTTGGTGTGGCTGGAAAACCATTTGGGATGGTTATTTCT	1620
Db	2658	ATTCCCATGAACTCGAGATAT	CAATTTGGTGTGGCTGGAAAATCATTTGGGATGGCTATTTCT	2717
Qy	1621	AAAGGTATCAACAGAAAAC	TTTGAAAAACAGGGTTATATCCCTCTTACAAAGTCCAGAG	1680
Db	2718	AAAGGTGTCAACAGGAAAT	TTGGAGGAGCGGGCTATATCCCTCTTACAAAGTTCGAGAG	2777
Qy	1681	AAGATGAAACAGTCAAGTAT	CCCATATCTCTGAAGCTGMAAATA	1727
Db	2778	AAGATGAAACGGTCAAGTAC	CCCAATATCTCTGAGGCTGAGAAATA	2824

Search completed: February 2, 2004, 16:22:55  
Job time : 6350.17 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 10:17:23 ; Search time 467.615 Seconds  
(without alignments)  
9975.371 Million cell updates/sec

Title: US-09-971-773-2  
Perfect score: 1728  
Sequence: 1 atcgggcgatgactggttc.....atcctgaagctgaaaaatag 1728

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 19Jun03:\*

1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	100.0	1728	24	ABK70032 Antibody productio
2	1584	91.7	2008	24	ABK70031 Antibody productio
3	1501.4	86.9	1759	24	ABA98809 Alpha1,6-fucosyl t
4	1501.4	86.9	2100	18	AAT76574 Human alpha 1-6 fu
5	1501.4	86.9	3291	23	ABV22468 Human prostate exp
6	1501.4	86.9	3291	23	ABV28283 Human prostate exp
7	1490	86.2	3007	22	AAH33315 Human colon cancer
8	1469.4	85.0	1728	18	AAT76573 Pig alpha 1-6 fuco

9	931	53.9	979	21	AAC63892 Rat FUT8 cDNA. Ra
10	931	53.9	979	24	ABK70037 Antibody productio
11	894.2	51.7	979	31	AAC63891 Chinese hamster FU
12	894.2	51.7	979	34	ABK70036 Antibody productio
13	876.2	50.7	1017	22	AAF87952 Human alpha 1-6 fu
14	604.6	35.0	699	22	AAF87953 Human alpha 1-6 fu
15	378.8	21.9	2761	23	ABLO4601 Drosophila melanog
16	236.4	17.2	394	20	AAV89049 EST clone CS100.
17	269.2	15.6	503	22	AAK11265 Human brain expres
18	269.2	15.6	503	22	AAI42879 Human brain expres
19	241.8	14.0	4682	23	ABL04600 Drosophila melanog
20	210.8	12.2	551	32	AAK11037 Human brain expres
21	210.8	12.2	551	32	AAI42659 Probe #11345 used
22	205.4	11.9	248	22	ABA47883 Human foetal cell
23	205.4	11.9	248	22	ABA65776 Human foetal liver
24	205.4	11.9	248	22	ABA32861 Probe #11327 for g
25	205.4	11.9	248	22	AAK14177 Human brain expres
26	205.4	11.9	248	22	AAK25142 Human brain expres
27	205.4	11.9	248	22	AAK39911 Human bone marrow
28	205.4	11.9	248	22	AAI20722 Probe #10655 for g
29	205.4	11.9	248	22	AAI45939 Probe #14625 used
30	205.4	11.9	248	22	AAI57183 Probe #25869 used
31	205.4	11.9	248	22	AAI06420 Probe #6411 used t
32	205.4	11.9	248	23	ABS39501 Human liver single
33	205.4	11.9	248	24	ABS14010 Human genome-deriv
34	188.6	10.9	9196	24	ABK70033 Antibody productio
35	147.4	8.5	481	25	ABT21901 Breast cancer mark
36	126.8	7.3	384	22	ABA42755 Human breast cell
37	126.8	7.3	384	22	ABA53183 Human foetal liver
38	126.8	7.3	384	22	ABA22955 Probe #1421 for ge
39	126.8	7.3	384	22	AAK01429 Human brain expres
40	126.8	7.3	384	22	AAK26877 Human bone marrow
41	126.8	7.3	384	22	AAI11508 Probe #1441 for ge
42	126.8	7.3	384	22	AAI32785 Probe #1471 used t
43	126.8	7.3	384	22	AAI01422 Probe #1413 used t
44	126.8	7.3	384	23	ABS26469 Human liver single
45	126.8	7.3	384	24	ABS01477 Human genome-deriv

ALIGNMENTS

RESULT 1

ID ABK70032 standard; cDNA; 1728 BP.  
AC ABK70032;  
15-JUL-2002 (first entry)  
DE Antibody production method related cDNA #2.

XX Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.

OS Mus musculus.

PN WO200231140-A1.

PD 18-APR-2002.

XX 05-OCT-2001; 2001WO-JP08804.

XX 06-OCT-2000; 2000JP-0308526.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX WPI, 2002-340182/37.

XX Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections  
XX  
PS Claim 32; Page 4-11; 314pp; Japanese.  
XX  
CC This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transfected with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The  
CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a nucleotide  
CC molecule used in the method of the invention.  
XX  
SQ Sequence 1728 BP; 529 A; 365 C; 400 G; 434 T; 0 other;  
Query Match 100.0%; Score 1728; DB 24; Length 1728;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCGGGCATGACGTGGTCTCGGGTGGGATATGCTCATCTTTTGGCCCTGGGGACC 60  
DB 1 ATGCGGGCATGACGTGGTCTCGGGTGGGATATGCTCATCTTTTGGCCCTGGGGACC 60  
QY 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCGAGATATGACCACTTCATCTCCAGC 120  
DB 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCGAGATATGACCACTTCATCTCCAGC 120  
QY 121 AGAGAACTCTCCAGATTTCTGCAAGCTTTGAAAGCTTTAAACAGCAAAATGAAGACTTG 180  
DB 121 AGAGAACTCTCCAGATTTCTGCAAGCTTTGAAAGCTTTAAACAGCAAAATGAAGACTTG 180  
QY 181 AGGCGAATGGCTGAGTCTCTCGAATACAGAGGCCCCATGACGGGAGCAGCTACA 240  
DB 181 AGGCGAATGGCTGAGTCTCTCGAATACAGAGGCCCCATGACGGGAGCAGCTACA 240  
QY 241 GGAAGAGTCCGCTGTTTGAAGAAACAGCTTTGTTAAGGGCAAGAAACAGATTTGAAATTTAC 300  
DB 241 GGAAGAGTCCGCTGTTTGAAGAAACAGCTTTGTTAAGGGCAAGAAACAGATTTGAAATTTAC 300  
QY 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360  
DB 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360  
QY 361 AATGAGCTAAAGAGCTCTGGTTTTCTTACAAAGGAACTGAAGAAATTAAGCATTTA 420  
DB 361 AATGAGCTAAAGAGCTCTGGTTTTCTTACAAAGGAACTGAAGAAATTAAGCATTTA 420  
QY 421 GAAGGAATGAATCTCAAGAGATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
DB 421 GAAGGAATGAATCTCAAGAGATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
QY 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGGAGCGGGATTTGGCGT 540  
DB 481 AGGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGGAGCGGGATTTGGCGT 540  
QY 541 GAAAGAGGCCCCAAGATCTGACAGAGTGGTTCGAGCGGAGATTAACATATCTCCAGAT 600  
DB 541 GAAAGAGGCCCCAAGATCTGACAGAGTGGTTCGAGCGGAGATTAACATATCTCCAGAT 600  
QY 601 CCTAAGGACTGACGAAAGCAGGAGCTGGTGTGTAAATCAATCAATAAAGGCTGTGGCTAT 660  
DB 601 CCTAAGGACTGACGAAAGCAGGAGCTGGTGTGTAAATCAATCAATAAAGGCTGTGGCTAT 660  
QY 661 GGTGTCTCAACTCCATCACTGCTCTGCTTTTCAATGATTTGTCACCCAGCGGAACA 720

Db 661 GGTGTCTCAACTCCATCACTGCTCTGCTTTTCAATGATTTGTCACCCAGCGGAACA 720  
QY 721 CTCAATCTTGGAAATCTCAGAAATGGCGCTATGCTACTCTGGTGGATCGGAGACTGTGTTAGA 780  
DB 721 CTCAATCTTGGAAATCTCAGAAATGGCGCTATGCTACTCTGGTGGATCGGAGACTGTGTTAGA 780  
QY 781 CCTGTAAAGTGAAGATGTACAGAGATCTGGCCCTCTCCACTGACACACTGTCAGGTGAA 840  
DB 781 CCTGTAAAGTGAAGATGTACAGAGATCTGGCCCTCTCCACTGACACACTGTCAGGTGAA 840  
QY 841 GTAAATGACAAAACATTTCAAGTGGTGGTCTCCCATTTGTAGACAGCTCCATCTCGG 900  
DB 841 GTAAATGACAAAACATTTCAAGTGGTGGTCTCCCATTTGTAGACAGCTCCATCTCGG 900  
QY 901 CCTCTTACTTACCACTGGCTTTCCAGAGACCTTGCAGACCGACTTCTTAAGAGTCCAT 960  
DB 901 CCTCTTACTTACCACTGGCTTTCCAGAGACCTTGCAGACCGACTTCTTAAGAGTCCAT 960  
QY 961 GGTGACCTGTCAGTGTGGTGGTGTCCAGTTTGTCAAATCTTGAATTCGTCACAACCT 1020  
DB 961 GGTGACCTGTCAGTGTGGTGGTGTCCAGTTTGTCAAATCTTGAATTCGTCACAACCT 1020  
QY 1021 TGGCTGGAAGAGGAATAGAGAGAGCCACCAAGAGCTTTGGCTTCAACATCTCCAGTTATT 1080  
DB 1021 TGGCTGGAAGAGGAATAGAGAGAGCCACCAAGAGCTTTGGCTTCAACATCTCCAGTTATT 1080  
QY 1081 GGAGTCCATGTTCAGACCCAGACAGACAAAGTGGGAAACAGAGCAGCTTCCACCCATCGAG 1140  
DB 1081 GGAGTCCATGTTCAGACCCAGACAGACAAAGTGGGAAACAGAGCAGCTTCCACCCATCGAG 1140  
QY 1141 GAGTACATGTTGACAGTTGAGAAACATTTTCAGCTTCTCGCAGCAGAGATGCAAGTGGAT 1200  
DB 1141 GAGTACATGTTGACAGTTGAGAAACATTTTCAGCTTCTCGCAGCAGAGATGCAAGTGGAT 1200  
QY 1201 AAAAAAGAGTATCTGGCTACCTGATGATCTACTTTGTTAAAGAGGAGGAAAGACAAG 1260  
DB 1201 AAAAAAGAGTATCTGGCTACCTGATGATCTACTTTGTTAAAGAGGAGGAAAGACAAG 1260  
QY 1261 TACTTCAATATGAAATTTATTTAGTGATACTCTATTTCTTGGTCAGCTGGACTACACAAT 1320  
DB 1261 TACTTCAATATGAAATTTATTTAGTGATACTCTATTTCTTGGTCAGCTGGACTACACAAT 1320  
QY 1321 CGGTACACAGAAATTCATTTCTGGGGTGTGATCTCTGATATACACTTTCTCTCAGGCT 1380  
DB 1321 CGGTACACAGAAATTCATTTCTGGGGTGTGATCTCTGATATACACTTTCTCTCAGGCT 1380  
QY 1381 GACTTTCTAGTGTGATCTTTTTCATCCAGCTCTGTCGGGTGCTTATGAAATCATGCA 1440  
DB 1381 GACTTTCTAGTGTGATCTTTTTCATCCAGCTCTGTCGGGTGCTTATGAAATCATGCA 1440  
QY 1441 ACCCTGCTATCTGATGCTCTGCGAACTTCCATTTCTTTGGATGACATCTACTATTTTGA 1500  
DB 1441 ACCCTGCTATCTGATGCTCTGCGAACTTCCATTTCTTTGGATGACATCTACTATTTTGA 1500  
QY 1501 GGCCAAATGCCCAATTCAGATTTGCTTTTATCTCACAACCTCTGAACTGAGAGGAA 1560  
DB 1501 GGCCAAATGCCCAATTCAGATTTGCTTTTATCTCACAACCTCTGAACTGAGAGGAA 1560  
QY 1561 ATTCCAATGGAACTGGAGATATCTTTGTTGGTGGCTGGAACCATTTGGATGTTATTTCT 1620  
DB 1561 ATTCCAATGGAACTGGAGATATCTTTGTTGGTGGCTGGAACCATTTGGATGTTATTTCT 1620  
QY 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCTCTCTCAAAAAGTCCGAGAG 1680  
DB 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCTCTCTCAAAAAGTCCGAGAG 1680  
QY 1681 AAGATAGAAACAGTCAAGATATCCACATATCTCTGAAGCTGAAAAATAG 1728  
DB 1681 AAGATAGAAACAGTCAAGATATCCACATATCTCTGAAGCTGAAAAATAG 1728

RESULT 2

ABK70031	standard; cDNA; 2008 BP.				
XX	AC	ABK70031;			
XX					
DT	(first entry)				
XX					
DE	Antibody production method related cDNA #1.				
XX					
KW	Antibody production; cytostatic; immunomodulator; vasotropic; virucide;				
KW	antibacterial; antiinflammatory; antiallergic; allergy; inflammation;				
KW	autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;				
KW	tumour; circulatory disease; infection; primer; ss.				
XX					
OS	Cricetulus griseus.				
XX					
PN	WO200231140-A1.				
XX					
PD	18-APR-2002.				
XX					
PF	05-OCT-2001; 2001WO-JP08904.				
XX					
PR	06-OCT-2000; 2000JP-0308526.				
XX					
XX	(KYOW ) KYOWA HAKKO KOGYO KK.				
PA					
XX					
PI	Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;				
PI	Hosaka E, Yamano K, Yamasaki M, Hanai N;				
DR	WPI; 2002-340182/37.				
XX					
PT	Cells producing antibody compositions including antibody fragments and				
PT	fusion proteins with Fc domain of antibody, useful for prevention or				
PT	treatment of cancer, immune diseases, circulatory diseases and				
PT	infections -				
XX					
PS	Claim 13; Page 1-3; 314pp; Japanese.				
XX					
CC	This invention relates to novel method for antibody production				
CC	comprising a Chinese hamster ovarian tissue-originated (CHO) cell				
CC	transferred with a gene encoding an antibody molecule for producing a				
CC	composition comprising an antibody molecule with an Fc domain bonded				
CC	to the N-glycoside linkage complex sugar chain. The produced antibody				
CC	compositions are drugs for prevention or treatment of diseases				
CC	accompanying tumour, allergy or inflammation, autoimmune diseases,				
CC	circulatory diseases, and viral and bacterial infections. The				
CC	antibodies can be stably produced using the method of the invention				
CC	with high binding activity and potency thus leading to high safety and				
CC	reduced side effects when applied alone or in combination with other				
CC	drugs for therapy. The present sequence represents a nucleotide				
CC	molecule used in the method of the invention.				
XX					
SQ	Sequence 2008 BP; 626 A; 425 C; 465 G; 492 T; 0 other;				
Query Match	91.7%; Score 1584; DB 24; Length 2008;				
Best Local Similarity	94.8%; Pred. No. 0;				
Matches 1638; Conservative	0; Mismatches 90; Indels 0; Gaps 0;				
QY	1 ATGCGGGATGAGTACTGGTTTCCTGGCGTTGGATTATGCTCATCTTTTTTGCCCTGGGGACC	60			
Dd	100 ATGCGGGATGAGTACTGGTTTCCTGGCGTTGGATTATGCTCATCTTTTTTGCCCTGGGGACC	159			
QY	61 TTGTGTTTTATATAGTGGTGCATTTGGTTCCAGATAATGACCACTGATCATTCCAGC	120			
Dd	160 TTATTGTTTTATATAGTGGTGCATTTGGTTCCAGATAATGACCACTGATCATTCTTAGC	219			
QY	121 AGAGAATCTCCAAGATTTCTTGAAGCTTGAACGCTTAAACAAGCAAAAATGAAGACTTG	180			
Dd	220 AGAGAATCTCCAAGATTTCTTGAAGCTTGAACGCTTAAACAAGCAAAAATGAAGACTTG	279			
QY	181 AGGCGAATGGCTGAGTCTCTCCGAATACCAGAGGCCCATTTGACAGGGGACAGCTACA	240			
Dd	280 AGGCGAATGGCTGAGTCTCTCCGAATACCAGAGGCCCATTTGATCAGGGGACAGCTACA	339			

```
QY 1321 CGGTACACAGAAATTCACCTTCGGGGTGTGATCCTGGATATACACTTTCTTCACAGGCT 1380
DB |||||||
QY 1420 CGATACACAGAAATTCACCTTCGGGGCGGTGATCCTGGATATACACTTTCTCTCCAGGCT 1479
DB |||||||
QY 1381 GACTTCTTAGTGTGTACTTTTTCATCCAGGCTGTGTCGGGTTGCTTATGAATCATGCAA 1440
DB |||||||
QY 1480 GACTTCTTAGTGTGTACTTTTTCATCCAGGCTGTGTCGGGTTGCTTATGAATCATGCAA 1539
DB |||||||
QY 1441 ACCCTGCATCTCGATCGCTCTCGGAATTCCTATTCCTTTGGATGACATCTACTATTTTGA 1500
DB |||||||
QY 1540 ACACCTGCATCTCGATCGCTCTCGGAATTCCTATTCCTTTAGATGACATCTACTATTTTGA 1599
DB |||||||
QY 1501 GGCCTAAATGCCCAATCAGATTCCTGTTTATCTCTCAAAACCTCGAACTGAAGAGAA 1560
DB |||||||
QY 1600 GGCCTAAATGCCCAATCAGATTCCTGTTTATCTCTCAAAACCTCGAACTGAAGAGAA 1659
DB |||||||
QY 1561 ATTCCAAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGGTTATCT 1620
DB |||||||
QY 1660 ATCCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGGTTATCT 1719
DB |||||||
QY 1621 AAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTTATATCTCTCAAAAGTCCGAGAG 1680
DB |||||||
QY 1720 AAAGGTATCAACAGAAAACTAGAAAAACAGGCTTTATATCTCTCAAAAGTCCGAGAG 1779
DB |||||||
QY 1681 AGATAGAAACAGTCAAGTATCCCATATATCTCTGAAGCTGAAAAATAG 1728
DB |||||||
QY 1780 AGATAGAAACAGTCAAGTATCCCATATATCTCTGAAGCTGAAAAATAG 1827
DB |||||||
```

## RESULT 3

```
ABA98809
ID ABA98809 standard; cDNA; 1759 BP.
XX AC ABA98809;
XX DT 07-MAY-2002 (first entry)
XX DE Alpha1,6-fucosyl transferase cDNA.
XX KW Plant; glycoprotein; alpha1,6-fucosyl transferase; alpha1,6-FT;
XX KW enzyme; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 17..1744
XX FT /*tag= a
XX FT /product= "Alpha1,6-fucosyl transferase"
XX PN JP2001333787-A.
XX PD 04-DEC-2001.
XX PF 06-MAR-2001; 2001JP-0062704.
XX PR 22-MAR-2000; 2000JP-0081059.
XX PA (TANI/) TANIGUCHI N.
XX PA (SEKI/) SEKI T.
XX PA (FUJI/) FUJIYAMA K.
XX DR WPI; 2002-158816/21.
XX DR P-PSDB; ABB08405.
XX PT A plant cell with an animal type sugar chain adding function, for the
XX PT preparation of a glycoprotein with an animal type sugar chain -
XX PS Example 1; Page 22-24; 38pp; Japanese.
XX PS The invention relates to a plant cell with an animal type sugar chain
XX CC adding function, created by transforming a tobacco cell with a gene
XX CC encoding an enzyme derived from an animal which can transfer a fucose
XX CC residue to the reductive end acetylglucosamine residue of a sugar chain.
```

CC The gene that is introduced into the plant cell encodes the enzyme  
CC alpha1,6-fucosyl transferase. The method of the invention is useful for  
CC the preparation of a glycoprotein having animal type sugar chain. The  
CC current sequence represents alpha1,6-fucosyl transferase cDNA.

SQ Sequence 1759 BP; 533 A; 367 C; 418 G; 441 T; 0 other;

Query Match 86.9%; Score 1501.4; DB 24; Length 1759;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 1 ATCGGGGCATCGACTGGTTCCTGGCTTGGATTATGCTCATTCTTTTGGCTGGGGACC 60
DB |||||||
QY 17 ATCGGGCATCGACTGGTTCCTGGCTTGGATTATGCTCATTCTTTTGGCTGGGGACC 76
DB |||||||
QY 61 TTGTTATTTATATAGTGGTTCATTTGGTTCGAGATAATGACCACTTCATCTCCAGC 120
DB |||||||
QY 77 TTGCTGTTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCACTTCATCTCCAGC 136
DB |||||||
QY 121 AGAGAACTCTCAAGATTCCTGCAAGCTTGAACGCTTAAACAGCAAAATGAGACTTG 180
DB |||||||
QY 137 CGAGAACTCTCAAGATTCCTGCAAGCTTGAACGCTTAAACAGCAAAATGAGACTTG 196
DB |||||||
QY 181 AGCGAAATGGCTGAGTCTCTCCGAATACCAAGAGCCCTTACCAAGGAGACACTACA 240
DB |||||||
QY 197 AGCGAAATGGCGGAATCTCTCCGAATACCAAGAGCCCTTATTTGATCGAGGCCACTATA 256
DB |||||||
QY 241 GGAAGAGTCCGTGTTTATAGAAAGAAACAGCTTCTTAAAGCCAAAGAAACAGATGAAAATTAC 300
DB |||||||
QY 257 GGAAGAGTACGCGTTTTAGAAAGAGAGCAGCTTCTTAAAGCCAAAGAAACAGATGAAAATTAC 316
DB |||||||
QY 301 AAGAAACAGCTAGAAATTCCTGGGAGAGATCATGAAATCTTAAAGAGGAGGATTGAA 360
DB |||||||
QY 317 AAGAAACAGACAGAAATTCCTGGGAGAGATCATGAAATCTTAAAGAGGAGGATTGAA 376
DB |||||||
QY 361 AATGAGCTAAAGAGCTCTGCTTTTCTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420
DB |||||||
QY 377 AATGAGCTAAAGAGCTCTGCTTTTCTCAAGAGTGAATTTGAAGAAATTTAAAGACTTTA 436
DB |||||||
QY 421 GAAGAAATGAATCCAAAGACATGCAGATGAAATTTCTTTGGATTTTAGGACACCATGAA 480
DB |||||||
QY 437 GAAGAAATGAATCCAAAGACATGCAGATGAAATTTCTTTGGATTTTAGGACATCATGAA 496
DB |||||||
QY 481 AGGCTATATGACAGATCTATATCTCTAGTCAACAGATGGAGCAGGGGATTTGGGCT 540
DB |||||||
QY 497 AGGCTATATATGACGGATCTATATCTCTAGTCAACAGATGGAGCAGGGGATTTGGGCG 556
DB |||||||
QY 541 GAAAAGAGGCCAAAGATCTCAGAGAGCTGCTCAGCGGAGAAATAACATATCTCCAGAAAT 600
DB |||||||
QY 557 GAAAAGAGGCCAAAGATCTCAGAGAGCTGCTCAGCGGAGAAATAACATATCTCCAGAAAT 616
DB |||||||
QY 601 CCTAAGGACTGCAGCAAGCCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 660
DB |||||||
QY 617 CCCAAGGACTGCAGCAAGCCAAAGCTGGTGTGTAAATAATCAACAAAGGCTGTGGCTAT 676
DB |||||||
QY 661 GGTGTCTAACTCCATCAGCTGCTCTACTGTTTTCATGATTGCTTATGGCACCCAGCGAACA 720
DB |||||||
QY 677 GGTGTCTAGCTCCATCATGCTGCTCTACTGTTTTCATGATTGCTTATGGCACCCAGCGAACA 736
DB |||||||
QY 721 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGAGACTGTGTTTGA 780
DB |||||||
QY 737 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGAGACTGTATTTAGG 796
DB |||||||
QY 781 CCTGTAGTGAGACATGTATCAGACAGATCTCGGCTCTCCACTGGACACTGGTCAAGTGA 840
DB |||||||
QY 797 CCTGTAGTGAGACATGTATCAGACAGATCTCGGCTCTCCACTGGACACTGGTCAAGTGA 856
DB |||||||
QY 841 GTAATGACAAAAACATTTCAAGTGGTTCAGCTCCCCATTTGTAGACAGCTCCATCTCGG 900
DB |||||||
QY 857 GTGAAGGACAAAAATGTTTCAAGTGGTTCAGCTTCCATTTGTAGACAGCTTCTTATCCCCGT 916
DB |||||||
QY 901 CCTCTTACTTACCACTGGCTGTTTCCAGAGACCTTTCAGACCCGACTCTTAAGAGTCCAT 960
DB |||||||
```

```
Db 917 CCTCCATATTTACCTTGGCTGTACCAAGACCTCGCAGATCGACTTGTACGAGTGCAT 976
Qy 961 GGTGACCTCGAGTGTGGTGTCCAGTTTGTCAATATCTGATTCGTCACCACT 1020
Db 977 GGTGACCTCGAGTGTGGTGTCTCAGTTTGTCAATATCTGATTCGTCACCACT 1036
Qy 1021 TGGCTGGAAGAAATAGAAGACCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1080
Db 1037 TGGCTAGAAAAGAAATAGAAGACCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1096
Qy 1081 GGAGTCCATGTGACAGCAGACAGAAAGTGGAGCAGAGCCTTCCACCCCATCGAG 1140
Db 1097 GGAGTCCATGTGACAGCAGACAGAAAGTGGAGCAGAGCCTTCCATCCCATTTGAA 1156
Qy 1141 GAGTACATGGTACAGTGTGAAGACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1200
Db 1157 GAGTACATGGTGCATGTTGAAGACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGCAT 1216
Qy 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTAAAGAGGCAAGCAAAAG 1260
Db 1217 AAAAAAGAGTATTTGGCCACAGATGACCTTCTTTTAAAGAGGCAAGCAAAAG 1276
Qy 1261 TACTCCAAATGAAATTTATGATGATACTCTATTTCTTGGTCACTGAGTACACAAAT 1320
Db 1277 TACCCCAATATGAATTTATGATGATACTCTATTTCTTGGTCACTGAGTGCACAAAT 1336
Qy 1321 CGGTACACAGAAATTTCACTTGGGCTGTGATCCTGGATATACATTTCTCTCACAGGCT 1380
Db 1337 CGATACACAGAAATTTCACTTGGGAGTGTATCTGGATATACATTTCTCTCACAGGCA 1396
Qy 1381 GACTTCTAGTGTGTACTTTTCACTCCAGGCTGTGCTGGTGTCTTATGAATCATGCAG 1440
Db 1397 GACTTCTAGTGTGTACTTTTCACTCCAGGCTGTGCTGGTGTCTTATGAATCATGCAG 1456
Qy 1441 ACCCTGCATCTGATGCTCTGCGAATTTCCATTTCTTGGATGACATCTACTATTTTGA 1500
Db 1457 ACATCTACATCTGATGCTCTGCGAATTTTCACTCCAGGCTGTGCTGGTGTCTTATTTGG 1516
Qy 1501 GGCCAAATGCCCAATCAGATTGCTTTATCTCTCAAACTCCAGCTGAGAGGAA 1560
Db 1517 GGCCAGAAATGCCCAATCAGATTGCTTTATGCTCAACACCCAGCTGAGAGGAA 1576
Qy 1561 ATTCCAATGGAACTCGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGGTTATTCT 1620
Db 1577 ATTCCAATGGAACTCGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGGTTATTCT 1636
Qy 1621 AAGGTATCAACAGAAAATTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680
Db 1637 AAGGTGTCAACAGAAAATTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1696
Qy 1681 AAGTAGAAGACAGTCAAGTATCCCATATATCTGAGCTGAGAAATA 1727
Db 1697 AAGTAGAAGACAGTCAAGTATCCCATATATCTGAGCTGAGAAATA 1743
```

```
RESULT 4
AAT76574
ID AAT76574 standard; cDNA to mRNA; 2100 BP.
XX
AC AAT76574;
XX
DT 05-MAR-1998 (first entry)
DE Human alpha 1-6 fucosyltransferase gene.
XX
KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;
KW guanosine diphosphate; sugar chain synthesis; modification; antibody;
KW GlcNAc; cancer diagnosis; ss.
XX
OS Homo sapiens.
XX
PH Key
FT CDS
198..1925
Location/Qualifiers
```

```
FT XX
PN PN
XX XX
PD 31-JUL-1997.
XX
PF 23-JAN-1997; 97WO-JP00171.
XX
PR 22-JUL-1996; 96JP-0192360.
PR 24-JAN-1996; 96JP-0010365.
PR 21-JUN-1996; 96JP-0161648.
PR 24-JUN-1996; 96JP-0162813.
XX
PA (TOYM ) TOYO BOSEKI KK.
XX
PI Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
XX
DR WPI; 1997-393690/36.
XX
PT P-PSDB; AAW22125.
XX
PT Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for
PT synthesis and modification of sugar chains and used as an antigen
XX for production of diagnostic antibodies
XX
PS Claim 18; Page 39-43; 61pp; Japanese.
XX
CC AAT76573 and AAT76574 represent the coding sequences for the pig and
CC human alpha 1-6 fucosyltransferases of the invention, respectively. The
CC encoded enzyme transfers fucose from guanosine diphosphate to the
CC 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:
CC 1-4GlcNAc-beta 1-6(GlcNAc-beta 1-2Man-alpha 1-3)Man-beta
CC 1-4GlcNAc-beta 1-4(GlcNAc-beta 1-4(Fuc-alpha 1-6)GlcNAc-R. It has
CC 1-2Man-alpha 1-3)Man-beta 1-4(GlcNAc-beta 1-4(Fuc-alpha 1-6)GlcNAc-R. It has
CC an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the
CC pH range 4-10 after 5 hours at 4 degrees C. The optimum working
CC temperature of the enzyme is 30-37 degrees C. A bivalent metal is not
CC required for activity of the enzyme, and the enzyme is not inhibited in
CC the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
CC modification of sugar chains, and as antigen for the production of
CC antibodies recognising the enzyme. The antibodies can be used for the
CC diagnosis of cancer and other diseases.
XX
SQ Sequence 2100 BP; 651 A; 436 C; 489 G; 524 T; 0 other;
```

```
Query Match 86.9%; Score 1501.4; DB 18; Length 2100;
Best Local Similarity 91.8%; Pred. No. 0;
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy 1 ATCGGGGCGATGGACTGGTTCCTCGGCTTGGATTATGCTCATTTCTTTTGGCGGGACC 60
Db 198 ATCGGGCGATGGACTGGTTCCTCGGCTTGGATTATGCTCATTTCTTTTGGCGGGACC 257
Qy 61 TTGTTATTTTATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCCAGC 120
Db 258 TTGCTGTTTTATAGTGGTTCATTTGGTTCGAGATATGACCCCTGATCCTCTAGC 317
Qy 121 AGAAGCTCTCCAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180
Db 318 CGAAGACTGTCCAGATTTCTGCGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 377
Qy 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGGCCCTTGTGACAGGGGACAGCTACA 240
Db 378 AGGCGAATGGCGGAATCTCTCGGATACCAAGGCCCTTGTGATCAGGGGCCAGCTATA 437
Qy 241 GGAAGAGTCCGTGTTTATAGAAAGACAGCTTGTAAAGGCCAAAGAACAGATGAAATTAC 300
Db 438 GGAAGAGTACCGTTTATAGAAAGACAGCTTGTAAAGGCCAAAGAACAGATGAAATTAC 497
Qy 301 AAGAAACAGCTAGAAATGCTCTGGGAGAGTATCATGAATCTTAAAGAGGAGGATTCGA 360
Db 498 AAGAAACAGACAGAAATGGTCTGGGAGAGATCATGAATCTTAAAGAGGAGGATTCGA 557
Qy 361 AATGAGCTAAAGAGCTCTGGTGTCTTCTCAAAAGCGAACTGAAGAAATTAAGCATTTA 420
```

Db 558 AATGAGCTAAAGAGCTCTGGTGTCTCTTCTTCTACAGCTGAATTAAGAACTTA 617  
Qy 421 GAAGGAATGAATCTCAAGACATCGAGATGAATCTTTGGATTTAGGACACCATGAA 480  
Db 618 GAAGGAATGAATCTCAAGACATCGAGATGAATCTTTGGATTTAGGACATGAA 677  
Qy 481 AGGTCTATCATGACAGATCTATCTACTACCTCAGTCAAAACAGATGGAGCGGATTTGGCGT 540  
Db 678 AGGTCTATAATGACGACTATATACTACTACCTCAGTCAAGACAGATGGAGCGGATTTGGCGG 737  
Qy 541 GAAAGAGGCCAAAGATCTGACAGAGCTGGTTCAGCGGAGATTAACATATCTCCAGAT 600  
Db 738 GAAAGAGGCCAAAGATCTGACAGAGCTGGTTCAGCGGAGATTAACATATCTCCAGAT 797  
Qy 601 CCTAAGGACTGACGAAAGCCAGGAGCTGGTGTCTTAACATCAATAAAGGCTGTGGCTAT 660  
Db 798 CCCAAGGACTGACGAAAGCCAAAGCTGGTGTCTTAACATCAATAAAGGCTGTGGCTAT 857  
Qy 661 GGTTCTCAACTCACTCACTGCTCTACTGTTCTTCTATGATTTGCTTATGGCACCCAGCGAACA 720  
Db 858 GGTCTGAGCTCCATCATGCTGTCTACTGCTTCATGATTTGATATGGCACCCAGCGAACA 917  
Qy 721 CTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGTTGATGGAGACCTGTGTTTGA 780  
Db 918 CTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGTTGATGGAGACCTGTATTTAGG 977  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGACACCTGTGAGTCAA 840  
Db 978 CCTGTAAGTGAGACATGTACAGACAGATCTGGCATCTCCACTGACACCTGTGAGTCAA 1037  
Qy 841 GTAAATGCAAAAAATTCAGTGTGAGCTGCCCATTTGTAGACAGCTCCATCTCTCGG 900  
Db 1038 GTGAGGACAAAAATTCAGTGTGAGCTGCCCATTTGTAGACAGCTCTCATCCCGT 1097  
Qy 901 CCTCTTACTTACCATCTGGCTGTTCCAGAGACCTTTGACAGACCGACTCTTAAAGTCCAT 960  
Db 1098 CCTCCATATTTACCTTGGCTGTACAGAGACCTTCGAGATCGAGTGTGACAGTGCAT 1157  
Qy 961 GGTGACCTGACGTGTGGTGTCCAGTGTTCCTCAATCTGATTCGTCACCACT 1020  
Db 1158 GGTGACCTGACGTGTGGTGTTCCTCAATCTGATTCGTCACCACT 1217  
Qy 1021 TGGCTGAAAAAGAAATAGAGAGCCACCAAGAGCTTTGGCTTCAAAACATCCAGTTAT 1080  
Db 1218 TGGCTGAAAAAGAAATAGAGAGCCACCAAGAGCTTTGGCTTCAAAACATCCAGTTAT 1277  
Qy 1081 GGAGTCCATGTGACAGGACAGACAAAGTGGAAACAGAGCGCTTCCACCCCATTCGAG 1140  
Db 1278 GGAGTCCATGTGACAGGACAGACAAAGTGGAAACAGAGCGCTTCCACCCCATTCGAG 1337  
Qy 1141 GAGTACATGTTACGTTGAAGACATTTTACGTTCTCGACGAGAGTGCAGTGGAT 1200  
Db 1338 GAGTACATGTTGATGTTGAAGACATTTTACGTTCTTCGACGAGAGTGCAGTGGAT 1397  
Qy 1201 AAAAAAGAGATATCTGGCTACTCATGATCTCTCTTTGTTAAAGAGGCGCAAGCAAG 1260  
Db 1398 AAAAAAGAGTATTTGGCCACAGATGACCTTCTTTTAAAGAGGCGCAAGCAAG 1457  
Qy 1261 TACTCCAAATTAATTAATTAATGATTAATCTATTTCTTGTGCTAGCTGGACATACAAAT 1320  
Db 1458 TACCCCAATTAATTAATTAATGATTAATCTATTTCTGCTAGCTGGACATGACAAAT 1517  
Qy 1321 CGGTACAGAAAAATCTCTGGGCTGTGATCTGATATACATTTCTCTCAGAGCT 1380  
Db 1518 CGATACAGAAAAATCTCTGGGAGTATCTGATATACATTTCTCTCTCAGGCA 1577  
Qy 1381 GACTTCTTAGTGTGATCTTTTCTATCCAGGCTGTCTCGGGTGTCTTATGAATCATGCAA 1440  
Db 1578 GACTTCTTAGTGTGATCTTTTCTATCCAGGCTGTCTCGAGTGTCTTATGAATCATGCAA 1637  
Qy 1441 ACCCTGATCTGATGCTCTGCGAACTTCATTTCTTTGGATGACATCTACTATTTTGA 1500

Db 1638 ACATACATCTGATGCTCTGCAAACTTCAATCTTTAGATGACATCTACTATTTTGG 1697  
Qy 1501 GGCCAAATGCCCAATCAGATTTGCTTTATCTCTCAAAACCTCGAACTGAAGAGAA 1560  
Db 1698 GGCCAGATGCCCAATCAGATTTGCTTTATCTCACCACCAACCCGAACTGCAGATGAA 1757  
Qy 1561 ATTCCAATGGAACTGGAGATATCATTTGGTGTGCTGGAACCAATTTGGATGTTATTTCT 1620  
Db 1758 ATTCCATGGAACTGGAGATATCATTTGGTGTGCTGGAATCATTTGGATGTTATTTCT 1817  
Qy 1621 AAAGGTATCAACAGAAACTTGGAAACAGGCTTATATCTCTCAAAAGTCCGAGAG 1680  
Db 1818 AAAGGTATCAACAGAAACTTGGAAAGGACGGGCTTATATCTCTCAAAAGTTCGAGAG 1877  
Qy 1681 AAAGTATGAAACAGTCAAGATATCCACATATCTCTGAAGCTGAAAAATA 1727  
Db 1878 AAGATAGAAAGGTCAGTACCTCCACATATCTCTGAGCTGAGAAATA 1924  
RESULT 5  
ABV22468  
ID ABV22468 standard; cDNA; 3291 BP.  
XX ABV22468;  
AC ABV22468;  
DT 13-SEP-2002 (first entry)  
XX Human prostate expression marker cDNA 22459.  
DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX Homo sapiens.  
OS WO200160860-A2.  
PN 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US05171.  
XX 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -  
Claim 1; Page 3914; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;  
CC

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX	Sequence	3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other;			
XX	Query Match	86.9%; Score 1501.4; DB 23; Length 3291;			
XX	Best Local Similarity	91.8%; Pred. No. 0;			
XX	Matches 1586; Conservative	0; Mismatches 141; Indels 0; Gaps 0;			
Qy	1	ATGCGGGATGACGCTGCTTCTGGCGTTGGATTATGCTCATCTCTTTTGGCTGGGGGACC	60		
Db	727	ATGCGGCATGACGCTGCTTCTGGCGTTGGATTATGCTCATCTCTTTTGGCTGGGGGACC	786		
Qy	61	TGTATTATTTATATAGTGGTCAATTTGGTTCGAGATAATGACCAACCTGATCCTCCAGC	120		
Db	787	TGTCTGTTTTATATAGTGGTCAATTTGGTTCGAGATAATGACCAACCTGATCCTCCAGC	846		
Qy	121	AGAGAACTCTCAAGATTTCTGCAAGCTTGAACGTTTAAACAGCAAAATGAAGACTTG	180		
Db	847	CGAGAACTGTCCAGATTTCTGCAAGCTTGAACGTTTAAACAGCAAAATGAAGACTTG	906		
Qy	181	AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTACA	240		
Db	907	AGGCGAATGGCGAATCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTATA	966		
Qy	241	GGAAGAGTCCGCTGTTTTAGAGAAACAGCTTTGTAAGGCCAAAGAACAGATTGAAAATTAC	300		
Db	967	GGAAGAGTACGCTGTTTTAGAGAGCAGCTTTGTAAGGCCAAAGAACAGATTGAAAATTAC	1026		
Qy	301	AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAGAGAGGAGATTGAA	360		
Db	1027	AAGAAACAGACCAAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAGAGAGGAGATTGAA	1086		
Qy	361	AATGAGCTAAAGAGCTCTGTTTTCTTACAAAGCGAACTGAAGAAATTAAGCATTTA	420		
Db	1087	AATGAGCTAAAGAGCTCTGTTTTCTTACAGAGTGAATTAAGAAATTAAGCACTTA	1146		
Qy	421	GAAGAAATGAATCTCAAGAACATGACAGATGAAATCTTTTGGATTATAGGACACCATGAA	480		
Db	1147	GAAGAAATGAATCTCAAGAACATGACAGATGAAATCTTTTGGATTATAGGACATCATGAA	1206		
Qy	481	AGGTCTATCATGACAGATCTATCTACCTCAGTCAAAACAGATGGAGCGGGATTGGCGT	540		
Db	1207	AGGTCTATAATGACGGATCTATCTACCTCAGTCAAAACAGATGGAGCGGGATTGGCGG	1266		
Qy	541	GAAGAAGGCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAAAT	600		
Db	1267	GAAGAAGGCCAAGATCTGACAGACTGGTTCAGCGGAGAAATAACATATCTCCAGAAAT	1326		
Qy	601	CCTAAGGAGCTGACGAAAGCCAGGAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT	660		
Db	1327	CCCAAGGAGCTGACGAAAGCCAAGAGCTGGTGTGTAAATAACAAGGCTGTGGCTAT	1386		
Qy	661	GGTTCTCACTTCATCAGTGGTCTACTGTTTCATGATTTGATTTGGCAACCCAGGAAACA	720		
Db	1387	GGCTGTCACTTCATCAGTGGTCTACTGTTTCATGATTTGATTTGGCAACCCAGGAAACA	1446		
Qy	721	CTCATCTTGGATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAGCTGTGTTTAGA	780		
Db	1447	CTCATCTTGGATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAGCTGTATTTAGG	1506		
Qy	781	CTGTAAAGTGAACATGTACACAGAGATCTGGCCCTCTCCACTTGGACACTGGTCAGGTGAA	840		
Db	1507	CTGTAAAGTGAACATGTACACAGAGATCTGGCATCTCCACTGGACACTGGTCAGGTGAA	1566		
Qy	841	GTAATGACAAAAATTCAAGTGTGAGCTCCCATTTGTAGACAGCCTCCATCCTCGG	900		
Db	1567	GTAAGGACAAAAATTCAAGTGTGAGCTCCCATTTGTAGACAGCTTTCATCCCGT	1626		
Qy	901	CCTCCTTACTTACACTGGCTGTTCCAGAGACCTTGCAGACCGACTCCTTAAGAGTCCAT	960		
Db	1627	CCTCCATATTTTACCTTGGCTGTACAGAGACCTTGCAGATCGACTTGTACGAGTGCAT	1686		
Qy	961	GGTGACCCCTGCAGTGTGGTGGTGTCCCGAGTTTGTCAAATACTTGTGCTGCCAACACT	1020		

Db	1687	GGTGACCCCTGCAGTGTGGTGGTGTCTCAGTTTGTCAAATACTTGATCCGCCACAGCT	1746		
Qy	1021	TGGCTGGAAGAGGAAATAGAAAGAACCCACCAAGAGCTTGGCTTCAAAATCCTCAGTTAT	1080		
Db	1747	TGGCTAGAAAGAAATAGAAAGAACCCACCAAGAGCTTGGCTTCAAAATCCTCAGTTAT	1806		
Qy	1081	GGAGTCCATGTGACAGCGCACAGCAAAAGTGGGAAACAGAGCAGCTTCCACCCCATCGAG	1140		
Db	1807	GGAGTCCATGTGACAGCGCACAGCAAAAGTGGGAAACAGAGCTTCCATCCCATCGAA	1866		
Qy	1141	GAGTACATGTGACAGCTTGAAGAAACATTTTCAGCTTCTCGCACGAGAAATGCAAGTGAAT	1200		
Db	1867	GAGTACATGTGACAGCTTGAAGAAACATTTTCAGCTTCTCGCACGAGAAATGCAAGTGAAT	1926		
Qy	1201	AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTAAAGAGGCGCAAGACAAAG	1260		
Db	1927	AAAAAAGAGTATATTTGGCCACAGATGACCCCTTCTTTTAAAGAGGCGCAAGACAAAG	1986		
Qy	1261	TACTCCAAATATGAATTTATTTAGTGAATCTATTTCTTGGTCAAGCTGGAGTACACAAT	1320		
Db	1987	TACCCCAATATGAATTTATTTAGTGAATCTATTTCTTGGTCAAGCTGGAGTACACAAT	2046		
Qy	1321	CGGTACACAGAAAATTTACCTTGGGGTGTGATCTCGATATACATTTCTCTCAGAGCT	1380		
Db	2047	CGATACACAGAAAATTTACCTTGGGGTGTGATCTCGATATACATTTCTCTCAGAGCT	2106		
Qy	1381	GACTTCTAGTGTCTACTTTTTCATCCAGCTCTGCGGGTGTCTTATGAATCATGCAA	1440		
Db	2107	GACTTCTAGTGTCTACTTTTTCATCCAGCTCTGCGGGTGTCTTATGAATCATGCAA	2166		
Qy	1441	ACCTTGCATCTGATGCTCTCGCAACCTTCCATTTCTTGGATGACATCTACTATTTTGA	1500		
Db	2167	ACATACATCTGATGCTCTCGCAACCTTCCATTTCTTGGATGACATCTACTATTTTGG	2226		
Qy	1501	GGCCAAATGCCCACATCAGATTGCTTTATCTCTCAAAACCTCGAACTGGAAGAGAA	1560		
Db	2227	GGCCAAATGCCCACATCAGATTGCTTTATCTCTCAAAACCTCGAACTGGAAGAGAA	2286		
Qy	1561	ATTCCAATGGAACTGAGATATCATTTGGTGTGGCTGGAACCATTTGGGATGGTTATCT	1620		
Db	2287	ATTCCAATGGAACTGAGATATCATTTGGTGTGGCTGGAACCATTTGGGATGGTTATCT	2346		
Qy	1621	AAAGTATCAACAGAAACTTTGGAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG	1680		
Db	2347	AAAGTATCAACAGAAACTTTGGAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG	2406		
Qy	1681	AAGATAGAAAACAGTCAAGTATCCACATATCTCTGAAAGCTGAAAATA	1727		
Db	2407	AAGATAGAAAACAGTCAAGTATCCACATATCTCTGAAAGCTGAAAATA	2453		
RESULT 6					
ABV28283	ID				
XX	ABV28283	standard; cDNA; 3291 BP.			
XX	AC	ABV28283;			
XX	DT	16-SEP-2002 (first entry)			
XX	XX	Human prostate expression marker cDNA 28274.			
DE	DE	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;			
XX	XX	pharmacogenomic marker; gene; ss.			
OS	OS	Homo sapiens.			
XX	XX	WO200160860-A2.			
XX	XX	23-AUG-2001.			
XX	XX	20-FEB-2001; 2001WO-US05171.			



PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PA Schlegel R, Endege WO, Monahan JB;  
PI WPI; 2001-662795/76.  
DR Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 5883; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other;  
SQ  
Query Match 86.9%; Score 1501.4; DB 23; Length 3291;  
Best Local Similarity 91.8%; Pred. No. 0;  
Matches 1586; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 1 ATGCGGCGATGACCTGGTCTCGGGCTGGGATGATGCTCATCTTTTGGCTGGGGGACC 60  
DB 727 ATGCGGCGATGACCTGGTCTCGGGCTGGGATGATGCTCATCTTTTGGCTGGGGGACC 786  
QY 61 TTGTTATTTTATAGGTGGTCACTTTGGTTCAGATATGACCCCTGATCCTCCAGC 120  
DB 787 TTGCTGTTTATATAGGTGGTCACTTTGGTTCAGATATGACCCCTGATCCTCTAGC 846  
QY 121 AGAGAACTCTCCAAGATCTTGCAGAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
DB 847 CGAGAACTCTCAAGATCTTGCAGAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 906  
QY 181 AGGCGAATGGCTGATCTCTCGAATACAGAGGCGCCCATTTGACGGGACAGCTACA 240  
DB 907 AGGCGAATGGCGAATCTCTCGGATACAGAGGCGCCCATTTGATCAGGGGCGAGCTATA 966  
QY 241 GGAAGAGTCCGTTTGAAGAACAGCTTTGTAAGGCGCAAGACAGATGAAATTAAC 300  
DB 967 GGAAGAGTCCGTTTGAAGAACAGCTTTGTAAGGCGCAAGACAGATGAAATTAAC 1026  
QY 301 AAGAAACAGCTTAGAAATGGTCTGGGAGAGATCATGAAATCTTAAAGAGGAGATTGAA 360  
DB 1027 AAGAAACAGACCAAGAAATGGTCTGGGAGAGATCATGAAATCTTGGGAGGAGATTGAA 1086  
QY 361 AATGAGCTAAAGAGCTCTGTTTCTTAAGGCGCAAGACAGATGAAATTAAGCAATTA 420  
DB 1087 AATGAGCTAAAGAGCTCTGTTTCTTACAGAGTGAATTAAGGCGCAAGACAGATGAA 1146  
QY 421 GAAGGAAATGAACTTCAAGACATCAGATGAAATCTTTTGGATTTTAGGACCATGAA 480  
DB 1147 GAAGGAAATGAACTTCAAGACATCAGATGAAATCTTTTGGATTTTAGGACATGAA 1206

QY 481 AGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGAGCAGGGATTGGCGT 540  
DB 1207 AGTCTATTAATGACGGATCTATACCTCAGTCAAGACAGATGAGCAGGGATTGGCGG 1266  
QY 541 GAAAAGAGGCGCAAGATCTGACAGAGCTGGTCAGGGGAGAAATAACATATCTCCAGAA 600  
DB 1267 GAAAAGAGGCGCAAGATCTGACAGAACTGGTCAGGGGAGAAATAACATATCTCCAGAA 1326  
QY 601 CCTAAGGACTGACAGAAAGCCAGGAAGCTGGTGTAAACATCAATAAAGGTGGCTTAT 660  
DB 1327 CCCAAGGACTGACAGAAAGCCAGGAAGCTGGTGTAAATAATCAACAAGGCTGGCTAT 1386  
QY 661 GGTGTCAACTCCATCAGCTGGTCTACTGTTTCATGATTGCTTATGGCACCAGGAGACA 720  
DB 1387 GGCTGTGAGCTCCATCATGCTGCTACTGCTTCATGATTGCTATGGCACCAGGAGACA 1446  
QY 721 CTCATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGCTGATGGAGACTGTGTTAGA 780  
DB 1447 CTCATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGCTGATGGAGACTGTATTAGG 1506  
QY 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGTCAGGTGAA 840  
DB 1507 CCTGTAAGTGAGACATGCACAGACAGATCTGGCATCTCCACTGGACACTGTCAGGTGAA 1566  
QY 841 GTAATGACAAAACATTTCAAGTGGTGGCTGCCCATTTGTCAGAGCTCCATCTCGG 900  
DB 1567 GTGAAGGACAAAATGTTCAAGTGGTGGCTGCCCATTTGTCAGAGCTCTTATCCCGCT 1626  
QY 901 CCTCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACCGACTCTTAAGAGTCCAT 960  
DB 1627 CCTCCATATTTACCCTTGGCTGTACAGAGACCTTGCAGATGCACTGTAGAGTGCAT 1686  
QY 961 GGTGACCTGTCAGTGTGGTGGCTGCCCATTTGTCAGAAATCTTGAATTCGTCACAACT 1020  
DB 1687 GGTGACCTGTCAGTGTGGTGGCTGCTCAGTTTGTCAAATCTTGTATCCGCCACAGCT 1746  
QY 1021 TGGCTGGAAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1080  
DB 1747 TGGCTAGAAAAGAAATAGAAAGACCCACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1806  
QY 1081 GGAGTCCATGTACAGCCACACAGCAAAAGTGGGAAACAGAGCAGCTTCCACCCCATCGAG 1140  
DB 1807 GGAGTCCATGTACAGCCACACAGCAAAAGTGGGAAACAGAGCTGCTTCCATCCCATGAA 1866  
QY 1141 GAGTACATGTGTACAGTTGAGAAACATTTTCAGCTTCTCGACGACAGAAATCAAGTGCAT 1200  
DB 1867 GAGTACATGTGTGTGAGAAACATTTTCAGCTTCTTTCAGCAGAGAAATGCAAGTGCAC 1926  
QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTACTTGTAAAGGAGGCAAGACAAG 1260  
DB 1927 AAAAAAGAGTATTTTGGCCACAGATGACCTTCTTTATTAAGGAGGCAAAACAAG 1986  
QY 1261 TACTCCATTAATGAATTTATTTAGTGAATCTATTTCTTGGTCACTGAGCTGACACAAT 1320  
DB 1987 TACCCCAATTAATGAATTTATTTAGTGAATCTATTTCTTGGTCACTGAGCTGACACAAT 2046  
QY 1321 CGGTACACAGAAAATTCACATTTGGGGTGTGATCTCGGATATACATTTCTCTCAGAGCT 1380  
DB 2047 CGATACACAGAAAATTCACATTTGGGAGTGTATCTGGATATACATTTCTCTCAGGCA 2106  
QY 1381 GACTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCA 1440  
DB 2107 GACTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTTGCTTATGAATCATGCA 2166  
QY 1441 ACCCTGATCTGATGCTCTCGCAACTTCCATTTCTTGGATGACATCTACTATTTTGA 1500  
DB 2167 ACATACATCTGATGCTCTCGCAACTTCCATTTCTTGGATGACATCTACTATTTTGG 2226  
QY 1501 GGCCAAAATGCCACAATCAGATTTGCTTTATCTCACAACCTCGAACTGAGAGGAA 1560  
DB 2227 GGCCAGAATGCCACAATCAAATTTGCCATTTATGCTACCAACCCCGAATCGCATGAA 2286  
QY 1561 ATTCAAATGGAACTCGGAGATATCATTTGGTGGCTGGAACCATTCGGATGGTTATCT 1620



Db 2287 ATCCCATGGAACCTGGAGATATCATTTGGTGGCTGGAATCATTTGGATGGCTATTTCT 2346  
Qy 1621 AAGGTATCAACAGAACTTGGAAACAGGCTTATATCCCTCTCAAGTCCGAGAG 1680  
Db 2347 AAAGGTGTCACAGAAATTTGGAAAGGCGGCTTATATCCCTCTCAAGATTCGAGAG 2406  
Qy 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAAATA 1727  
Db 2407 AAGATAGAAACGGTCNAGTACCCACATATCTCTGAGCTGAGAAATA 2453

RESULT 7  
AAH33315  
ID AAH33315 standard; cDNA; 3007 BP.  
XX  
AC AAH33315;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:371.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; chromosome 14; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
WPI; 2001-235357/24.  
DR P-PSDB; AAG73884.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
XX Claim 1; Page 2480-2481; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate p  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the patients own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated Ps,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
XX and AAH77789 represent sequences used in the exemplification of the  
XX present invention.  
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
XX missing at time of publication, meaning no sequences are present for  
XX SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
XX Sequence 3007 BP; 900 A; 597 C; 668 G; 834 T; 8 other;  
XX  
XX Query Match 86.2%; Score 1490; DB 22; Length 3007;  
XX Best Local Similarity 91.7%; Pred. No. 0;  
XX Matches 1585; Conservative 1; Mismatches 141; Indels 1; Gaps 1;

Qy 1 ATCGGGCATGAGTGGTCTCTGGCGTTGGATTATGCTCATTTTGGCTGGGGACC 60  
Db 420 ATCGGGCATGAGTGGTCTCTGGCGTTGGATTATGCTCATTTTGGCTGGGGACC 479  
Qy 61 TTGTTATTTTATATAGTGGTGTCTTTGGTTCGAGATTAATGACACCCCTGATCATCCAGC 120  
Db 480 TTGCTGTTTTATATAGTGGTGTCTTTGGTTCGAGATTAATGACCATCTCTGATCACTTAGC 539  
Qy 121 AGAGAACTCTCCAAAGATTCTTGGAAAGCTTGAACGCTTAAAAACAGCAAAATGAAGACTTG 180  
Db 540 CGAGAACTGTCCAAGATTCTGGCAAGCTTGAACGCTTAAAAACAGCAAAATGAAGACTTG 599  
Qy 181 AGCGAATGGCTG-AGTCTCTCTCCGAATACCCAGAGGCCCCCATTTGACAGGGGAGCAGCTTAC 239  
Db 600 AGCGAATGGCGGAAATCTCTCCGGATACCCAGAGGCCCCATTGATCAGGGGCGCAGCTAT 659  
Qy 240 AGGAAGAGTCCGTGTTTTGAAGAAACAGCTTTTAAAGGCCAAAGAACAGATTTGAAATTA 299  
Db 660 AGGAAGAGTACGGCTTTTTAGAAAGAGCAGCTTTTAAAGGCCAAAGAACAGATTTGAAATTA 719  
Qy 300 CAAGAAACAGCTAGAAATGCTCTGGGGAAGGATCATGAAATCTTAAGAAAGGAGATTTGA 359  
Db 720 CAAGAAACAGACCCAGAAATGCTCTGGGGAAGGATCATGAAATCTCTGAGGAGGAGATTTGA 779  
Qy 360 AAATGGAGCTAAAGAGCTCTGGTTTTTTTCTACAAAGCGAACTGAAGAAATTTAAAGCAATTT 419  
Db 780 AAATGGAGCTAAAGAGCTCTGGTTTTTTTCTACAGAGTGAATTTGAAGAAATTTAAAGAACTT 839  
Qy 420 AGAAGAAATGAATCTCAAAAGACATGAGATGAATTTCTTTTGGATTAGGACACCATGA 479  
Db 840 AGAAGAAATGAATCTCAAAAGACATGAGATGAATTTCTTTTGGATTAGGACATCATGA 899  
Qy 480 AAGTCTTATCATGACAGATCTATATCTACTACCTCAGTCACAGATGGAGCAGGATTTGGCG 539  
Db 900 AAGTCTTATTAATGACATGATCTATATCTACTACCTCAGTCACAGATGGAGCAGGATTTGGCG 959  
Qy 540 TGAAGAAAGAGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAA 599  
Db 960 GGAAGAAAGAGCGCAAGATCTGACAGAACTGGTTCAGCGGAGAAATACATATCTCCAGAA 1019  
Qy 600 TCCTAAGAGCTGACAGTCAAGCGCAGAGAGTGGTGTGTAACATCAATTAAGCGCTGTGGCTA 659  
Db 1020 TCCCAAGAGCTGACAGTCAAGCGCAGAGAGTGGTGTGTAACATCAATTAAGCGCTGTGGCTA 1079  
Qy 660 TGGTTGTTCAACTCCATCAAGCTGGTCTACTGTTTTCATGATTTGCTTATGCGACCCAGCGAAC 719  
Db 1080 TGGCTGTGAGCTCCATCATGTTGGTCTACTGCTTTCATGATTTGCGACCCAGCGAAC 1139  
Qy 720 ACTCATCTTGGAAATCAAGAAATTTGGCGCTATGCTACTGTTGGATGGGAGACTGTGTTTATG 779  
Db 1140 ACTCATCTTGGAAATCAAGAAATTTGGCGCTATGCTACTGTTGGATGGGAGACTGTATTTATG 1199  
Qy 780 ACCTGTAAGTGAGACATGTACAGACAGATCTGGGCTCTCCACTGAGACACTGGTCAGGTGA 839  
Db 1200 GCCTGTAAGTGAGACATGTACAGACAGATCTGGGCTCTCCACTGAGACACTGGTCAGGTGA 1259  
Qy 840 AGTAAATGACAAAGAAATTTCAAGTGGTGGAGCTCCCATTTGTAGACAGCTTCCATCTCTCG 899  
Db 1260 AGTAAAGGACAAAGAAATTTCAAGTGGTGGAGCTTCCCATTTGTAGACAGCTTTCATCTCCCG 1319  
Qy 900 GCCTCTCTTACTTACCACCTGGCTGTTTCCAGAAAGACCTTGCAGACCGACTCTCTAAGAGTCCA 959  
Db 1320 TCCTCCATATTTTACCCTTGGCTGTACCAAGAAAGACTTCGCAGATGCATTTGTACAGGTGA 1379  
Qy 960 TGGTGACCCCTGCAGTGTGGTGGTGTCCCGATTTGTCAAATATCTTGAATTCGTCCCAACACC 1019  
Db 1380 TGGTGACCCCTGCAGTGTGGTGGTGTCTCAGTTTGTCAAATATCTTGAATTCGTCCCAACACC 1439  
Qy 1020 TTGGCTGGAAGAAAGAAATAGAAAGGACCAAGAAAGCTTGGCTTCAAGACATCCAGTTAT 1079  
Db 1440 TTGGCTGGAAGAAAGAAATAGAAAGGACCAAGAAAGCTTGGCTTCAAGACATCCAGTTAT 1499

1080	Qy	TGGAGTCCATGTCAGACGCAAGCTGGGNA	1139
1080	Qy	TGGAGTCCATGTCAGACGCAAGCTGGGNA	1139
1500	Db	TGGAGTCCATGTCAGACGCAAGCTGGGNA	1559
1500	Db	TGGAGTCCATGTCAGACGCAAGCTGGGNA	1559
1140	Qy	GGAGTACATGGTACACGCTTGAAGAA	1199
1140	Qy	GGAGTACATGGTACACGCTTGAAGAA	1199
1560	Db	AGAGTACATGGTGCATGTTGAGNACAT	1619
1560	Db	AGAGTACATGGTGCATGTTGAGNACAT	1619
1200	Qy	TAAAAAAGAGTATATCTGGCTACTGAT	1259
1200	Qy	TAAAAAAGAGTATATCTGGCTACTGAT	1259
1620	Db	CAAAAAAGAGTGTATTTGGCCACAGAT	1679
1620	Db	CAAAAAAGAGTGTATTTGGCCACAGAT	1679
1260	Qy	GTACTCCAAATTATGAATTTATTAGTGA	1319
1260	Qy	GTACTCCAAATTATGAATTTATTAGTGA	1319
1680	Db	GTACCCCAATTTATGAATTTATTAGTGA	1739
1680	Db	GTACCCCAATTTATGAATTTATTAGTGA	1739
1320	Qy	TCGGTACACAGAAAATTCACCTTCGGG	1379
1320	Qy	TCGGTACACAGAAAATTCACCTTCGGG	1379
1740	Db	TCGATACACAGAAAATTCACCTTCGGG	1799
1740	Db	TCGATACACAGAAAATTCACCTTCGGG	1799
1380	Qy	TGACTTTCTAGTGTACTTTTTCATCCCA	1439
1380	Qy	TGACTTTCTAGTGTACTTTTTCATCCCA	1439
1800	Db	AGACTTCTAGTGTACTTTTTCATCCCA	1859
1800	Db	AGACTTCTAGTGTACTTTTTCATCCCA	1859
1440	Qy	AACCTTCATCTTGATGCTCTCGGAAC	1499
1440	Qy	AACCTTCATCTTGATGCTCTCGGAAC	1499
1860	Db	AACACTACATCTTGATGCTCTCGAACC	1919
1860	Db	AACACTACATCTTGATGCTCTCGAACC	1919
1500	Qy	AGGCCAAAATGCCCAACATCAGATTGCT	1559
1500	Qy	AGGCCAAAATGCCCAACATCAGATTGCT	1559
1920	Db	GGGCCAAGATGCCCAACATCAAAATGG	1979
1920	Db	GGGCCAAGATGCCCAACATCAAAATGG	1979
1560	Qy	AATTCCAATGGAACTTGAGATATCAT	1619
1560	Qy	AATTCCAATGGAACTTGAGATATCAT	1619
1980	Db	AATTTCCCATGGAACTTGAGATATCAT	2039
1980	Db	AATTTCCCATGGAACTTGAGATATCAT	2039
1620	Qy	TAAAGGTATCAACAGAAAACTTGGAAAA	1679
1620	Qy	TAAAGGTATCAACAGAAAACTTGGAAAA	1679
2040	Db	TAAAGGTGTCAACAGGAAAATTTGGNAG	2099
2040	Db	TAAAGGTGTCAACAGGAAAATTTGGNAG	2099
1680	Qy	GAGATAGAAACAGTCAAGTATCCCAAT	1727
1680	Qy	GAGATAGAAACAGTCAAGTATCCCAAT	1727
2100	Db	GAGATAGAAACCGTCAAGTACCCCAAT	2147
2100	Db	GAGATAGAAACCGTCAAGTACCCCAAT	2147

## RESULT 8

AAT76573  
ID AAT76573 standard: cDNA to mRNA: 1728 BP.

AA  
AC AAT76573;

DT 05-MAR-1998 (first entry)

DE Piq alpha 1-6 fucosyltransferase gene.

KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GlcNAc; cancer diagnosis; ss.

OS Sus scrofa.

Key	Location/Qualifiers
CDS	1..1728
FT	/*tag= a

PN WO9727303-A1.

31-JUL-1997.

23-JAN-1997; 97WO-JP00171.

XX  
PR 22-JUL-1996; 96JP-0192260.

FN Z4-CHAN-1996; 96JF-0010363.

21-JUN-1996; 96JP-0161648.  
24-JUN-1996; 96JP-0162813.  
(TOYM) TOYO BOSEKI KK.  
Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
WPI; 1997-393690/36.  
P-PSDB; AAW22124.  
Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for  
synthesis and modification of sugar chains and used as an antigen  
for production of diagnostic antibodies  
Claim 5; Page 30-34; 61pp; Japanese.  
AAW7573 and AAW7574 represent the coding sequences for the pig and  
human alpha 1-6 fucosyltransferases of the invention, respectively. The  
encoded enzyme transfers fucose from guanosine diphosphate to the  
6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:  
(GlcNAceta 1-2Manalpa 1-6) (GlcNAceta 1-2Manalpa 1-3)Manbeta  
1-4GlcNAceta 1-4GlcNAc-R to give (GlcNAceta 1-2Manalpa 1-6) (GlcNAceta  
1-2Manalpa 1-3)Manbeta 1-4GlcNAceta 1-4GlcNAceta 1-4(Fucalpa 1-6)GlcNAc-R. It has  
an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the  
pH range 4-10 after 5 hours at 4 degrees C. A bivalent metal is not  
required for activity of the enzyme, and the enzyme is not inhibited in  
the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
modification of sugar chains, and as antigen for the production of  
antibodies recognising the enzyme. The antibodies can be used for the  
diagnosis of cancer and other diseases.  
Sequence 1728 BP; 521 A; 362 C; 419 G; 426 T; 0 other;

Query Match	85.0%;	Score 1469.4;	DB 18;	Length 1728;
Best Local Similarity	90.7%;	Pred. No. 0;		
Matches 1566;	Conservative	0;	Mismatches 161;	Indels 0; Gaps 0;
Qy	1	ATCGGGCATGGACTGGTTCTCGGCGTTGGATTATGCTCATTTCTTTTTCGCTGGGGGACC	60	
Db	1	ATCGGGCATGGACTGGTTCTGGCGTTGGATTATGCTCATTTCTTTTTCGCTGGGGGACC	60	
Qy	61	TTGCTTATTTTATATAGGTGGTCAATTTGGTTCGAGATAATGACCAACCCCTGATCATCTCCAGC	120	
Db	61	TTGCTATTTTACATAGGTGGTCACTTTGGTACGAGATAATGACCACTCTGATCACTCTAGC	120	
Qy	121	AGAGAACTCTCCAAAGATTCTTGCAAAAGCTTGAACGCTTAAACAGCAGCAAAATGAGAACTTGG	180	
Db	121	CGAGAACTGTGCCAAGATTTTGGCAAAAGCTGGAAACGCTTAAAAACACAAAATGAAGAACTTG	180	
Qy	181	AGSCGAATGCGTGAGTCTCTCCGAATACAGAAAGGCCCATTTGACCAAGGGGACAGACTACA	240	
Db	181	AGGAGAATGCGTGAATCTCTCCGAATACCAAGAAAGGCCCATTTGATCAGGGGCCAGCTTCA	240	
Qy	241	GGAAAGAGTCGCTGTTTTAGAGAACAACAGCTTGTTTAAAGGCCAAAGAAACAGATTGAAAAATTAC	300	
Db	241	GGAAAGAGTCGCTGTTTAGAGAGCAATTTATGAAGGCCAAAGAAACAGATTGAAAAATTAT	300	
Qy	301	AAGAAAACAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAAGGAGGATTGAA	360	
Db	301	AAGAAAACAACATAAAATGGTCCAGGGAAGGATCATGAAATCTTAAAGGAGGAGGATTGAA	360	
Qy	361	AATGGAGCTTAAGAGCTCTGGTTTTTTTCTACAAGCGAACTGAAGAAATTAAGACATTTA	420	
Db	361	AATGGAGCTTAAGAGCTCTGGTTTTTTTCTACAAGAGTGTGAAGAAATTAAGAAATTTA	420	
Qy	421	GAAGGAAATGAACCTCCAAAGACATGCAGATGAAATTTCTTTTGGATTTTAGGACACCATGAA	480	
Db	421	GAAGGAAATGAACCTCCAAAGACATGCAGATGAATTTCTATCAGATTTGGGACATCATGAA	480	
Qy	481	AGGCTCTATCATGACAGATCTATACCTACGTCCTCAACAGATGGAGCAGGGGATTTGGCGT	540	
Db	481	AGGCTCTATATGACGAGATCTATACCTACCTCAGTCAACAGATGGGACAGGATTTGGCGT	540	

541 GAAAAGAGGCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAAAT 600  
Db |||||  
541 GAAAAGAGGCCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATAACATATCTCCAGAAAT 600  
Qy CCTAAGGACTGACGACAAAGCCAGGAGCTGGTGTGTGTAAACATCAATAAAGGCTGGCTAT 660  
Db |||||  
601 CCTAAGGACTGACGACAAAGCCAGGAGCTGGTGTGTGTAAACATCAATAAAGGCTGGCTAT 660  
Qy CCCAAGGACTGACGACAAAGCCAGGAGCTGGTGTGTGTAAATATCAACAAAGGCTGGCTAT 660  
Db |||||  
661 GGTGTCAACTCCATCACTGCTGCTTACTGTTTCATGATGCTTATGGCCACCAGGACAA 720  
Qy |||||  
661 GGTGTCAACTCCATCACTGCTGCTTACTGTTTCATGATGCTTATGGCCACCAGGACAA 720  
Db |||||  
721 CTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGCTGGATGGGAGCTGTGTTTGA 780  
Qy |||||  
721 CTCGCTTGGAAATCTCAGAAATGGCGCTATGCTACTGCTGGATGGGAGCTGTGTTTGA 780  
Db |||||  
781 CCTGTAAGTGACATGTACAGACAGATCTGGCCTCTCCACTGGGACACTGTCAGGTGAA 840  
Qy |||||  
781 CCTGTAAGTGACATGTACAGACAGATCTGGCCTCTCCACTGGGACACTGTCAGGTGAA 840  
Db |||||  
841 GTAATGACAAACATCTCAAGTGTGAGCTGCCATTTAGAGAGCTTCCATCTCTCGG 900  
Qy |||||  
841 GTAAGGACAAACATCTCAAGTGTGAGCTGCCATTTAGAGAGCTTCCATCTCTCGT 900  
Db |||||  
901 CCTCTTACTTACCACTGCTGCTTCCAGAGACCTTGCAGACCGACTCCTAAGAGTCCAT 960  
Qy |||||  
901 CCTCTTACTTACCACTGCTGCTTCCAGAGACCTTGCAGACCGACTCCTAAGAGTCCAT 960  
Db |||||  
961 GGTGACCTGAGTGTGGTGTGCTCCAGATTTTGTCAATATCTGATTCGTCACAACT 1020  
Qy |||||  
961 GGTGACCTGAGTGTGGTGTGCTCCAGATTTTGTCAATATCTGATTCGTCACAACT 1020  
Db |||||  
1021 TGGCTGGAAGGAAATGAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Qy |||||  
1021 TGGCTGGAAGGAAATGAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Db |||||  
1081 GGAGTCCATGTGACGACGACAGACAAAGTGGGAGAGAGCTTCCACCCCATCGAG 1140  
Qy |||||  
1081 GGAGTCCATGTGACGACGACAGACAAAGTGGGAGAGAGCTTCCACCCCATCGAG 1140  
Db |||||  
1141 GAGTACATGTTGACGTTGAGAGCAATTTTCAGCTTCTCGACGAGAGTCAAGTGGAT 1200  
Qy |||||  
1141 GAGTACATGTTGACGTTGAGAGCAATTTTCAGCTTCTCGACGAGAGTCAAGTGGAT 1200  
Db |||||  
1201 AAAAAGAGTATATCTGGCTACTGATGATCTCTTTGTTAAAGGAGGCAAGCAAG 1260  
Qy |||||  
1201 AAAAAGAGTATATCTGGCTACTGATGATCTCTTTGTTAAAGGAGGCAAGCAAG 1260  
Db |||||  
1261 TACTCCAATTATGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
Qy |||||  
1261 TACTCCAATTATGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
Db |||||  
1321 CGGTACACAGAAATTCATCTCGGGGTGTGATCCCTGGATATACATTTCTCTCAGAGCT 1380  
Qy |||||  
1321 CGGTACACAGAAATTCATCTCGGGGTGTGATCCCTGGATATACATTTCTCTCAGAGCT 1380  
Db |||||  
1381 GACTTTCTAGTGTGATTTTATCTCCAGGCTGCTCGGGTGTGATTTGAAATCATGAA 1440  
Qy |||||  
1381 GACTTTCTAGTGTGATTTTATCTCCAGGCTGCTCGGGTGTGATTTGAAATCATGAA 1440  
Db |||||  
1441 ACCCTGATCTGATGCTCTCGGAACTTCCATTTCTTTGGATGATCTACTATTTTGA 1500  
Qy |||||  
1441 ACCCTGATCTGATGCTCTCGGAACTTCCATTTCTTTGGATGATCTACTATTTTGA 1500  
Db |||||  
1501 GGGCAAAATGCCCAATCAGATTTGCTGTTTATCTCTCAAAACCTTCGAAGAGGAA 1560  
Qy |||||  
1501 GGGCAAAATGCCCAATCAGATTTGCTGTTTATCTCTCAAAACCTTCGAAGAGGAA 1560  
Db |||||  
1561 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCAATTTGGATGTTTCT 1620  
Qy |||||  
1561 ATTCCAATGGAACTGGAGATATTTATTTGGTGGCTGGAAATCACTGGGATGCTTCT 1620  
Db |||||

Qy 1621 AAGGTATACAGAAAACCTTGGAAAACAGGCTTATATCCCTCTACAAAGTCCGAGAG 1680  
Db |||||  
1621 AAGGTATACAGAAAACCTTGGAAAACAGGCTTATATCCCTCTACAAAGTCCGAGAG 1680  
Qy 1681 AAGTATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAATA 1727  
Db |||||  
1681 AAGTATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAATA 1727  
Qy 1681 AAGTATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAAATA 1727  
Db |||||

RESULT 9  
AAC63892  
ID AAC63892 standard; cDNA; 979 BP.  
XX  
AC AAC63892;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Rat FUT8 cDNA.  
XX  
KW Immunologically functional molecule; immune system; immunomodulation;  
KW glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;  
KW viral infection; bacterial infection; allergy; autoimmune disease;  
KW inflammation; antibody; rat; FUT8; ss.  
XX  
OS Rattus sp.  
XX  
PN WO200061739-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 07-APR-2000; 2000WO-JP02260.  
XX  
PR 09-APR-1999; 99JP-0103158.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;  
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;  
XX  
DR WPI; 2000-665129/64.  
XX  
PT Control of activity of antibodies and other immunologically functional  
PT molecules by addition or removal of fucose from sugar chain for  
PT diagnosis and treatment of cancer, allergy and other diseases -  
XX  
PS Example 8; Page 73-74; 81pp; Japanese.  
XX  
CC The invention relates to a method for controlling the activity of an  
CC immunologically functional molecule (e.g., an antibody) where the  
CC control is effected by the presence or absence of fucose bound to an  
CC N-acetylglucosamine residue at the reducing end of the sugar chain on  
CC the immunologically functional molecule. The invention also relates to  
CC methods for the diagnosis, prevention or treatment of diseases which  
CC involve the modified immunologically functional molecule, and agents  
CC which stimulate the activity of an immunologically functional molecule.  
CC The methods of the invention are used for the diagnosis, treatment and  
CC prevention of a broad range of diseases including cancer, circulatory  
CC disease, viral or bacterial infection, allergy, autoimmune disease and  
CC inflammation. The present sequence represents a rat FUT8 cDNA isolated  
CC in an exemplification of the invention.  
XX  
SQ Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;  
Query Match 53.9%; Score 931; DB 21; Length 979;  
Best Local Similarity 96.9%; Pred. No. 1.4e-272;  
Matches 949; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 720 ACTCATCTTGAATCTCAGAAATTCGCGCTATGCTACTGCTGGATGGGAGACTGTGTTAG 779  
Db |||||  
1 ACTCATCTTGAATCTCAGAAATTCGCGCTATGCTACTGCTGGATGGGAGACTGTGTTAG 60  
Qy 780 ACCTGTAGTACAGACATGTACAGACAGATCTGGGCTCTCCACTGGACACTGTCAGGTGA 839  
Db |||||

```
Db 61 ACTGTAAAGTGAGACATGACACAGACAGATCTGGCCCTCTCCACTGGACACTGGTTCAGGTGA 120
Qy 840 AGTAATAAGCAAAACATTTCAAGTGGTGGAGTCCCTCCCTTGTAGACAGCTCCATCTCTCG 899
Db 121 AGTGAATGACAAATATTTCAAGTGGTGGAGTCCCTCCCTTGTAGACAGCTTCATCTCG 180
Qy 900 GCCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGGAGACCGACTCTCTAAAGAGTCCA 959
Db 181 GCCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGGAGACTCGTAAAGAGTCCA 240
Qy 960 TGGTGACCTGCAGTGGTGGTGGTCCAGTTTCTCAATCTTCAATCTTCAATCTTCAATCTTCA 1019
Db 241 TGGTGATCTGCAGTGGTGGTGGTCCAGTTTCTCAATCTTCAATCTTCAATCTTCAATCTTCA 300
Qy 1020 TTGGCTGGAAGAAATAGAGAGAGCCACCAAGAGCTTGGCTTCAAACTATCCAGTTAT 1079
Db 301 TTGGCTGAGAAAGAAATAGAGAGAGCCACCAAGAGCTTGGCTTCAAACTATCCAGTTAT 360
Qy 1080 TGGAGTCCATGTACAGCGCACAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGA 1139
Db 361 TGGAGTCCATGTACAGCGCACAGACAAAGTGGGAACAGAGCAGCTTCCATCCCATCGA 420
Qy 1140 GGAGTACATGGTACAGTTGAAGACATTTTCAGCTTCTCGACCGACAGATGCCAGTCCA 1199
Db 421 AGAGTACATGGTACATGTGAAGACATTTTCAGCTTCTCGACCGACAGATGCCAGTCCA 480
Qy 1200 TAAAGAAAGAGTATCTGGCTACTGTAGTATCTTCTTAAAGAGGAGGACAAAGACAAA 1259
Db 481 TAAAGAAAGAGTATCTGGCTACTGTAGTATCTTCTTAAAGAGGAGGACAAAGACAAA 540
Qy 1260 GTACTCCAAATTAAGTAATTTATAGTAACATCTATTTCTGGTGGTGGAGTACACAA 1319
Db 541 GTACTCCAAATTAAGTAATTTATAGTAACATCTATTTCTGGTGGTGGAGTACACAA 600
Qy 1320 TGGGTACACAGAAATTCACCTCGGGGTGTGATCTCGATATACATCTTCTCTACAGGC 1379
Db 601 TCGGTACACAGAAATTCACCTCGGGGTGTGATCTCGATATACATCTTCTCTACAGGC 660
Qy 1380 TGACTTCTTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTGTGTTATGAATCATGCA 1439
Db 661 TGACTTCTTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTGTGTTATGAATCATGCA 720
Qy 1440 AACCTGCATCTGTAGTCTCGAACTTCCATTTCTTGGATGACATCTACTATTTGG 1499
Db 721 AACCTGCATCTGTAGTCTCGAACTTCCATTTCTTGGATGACATCTACTATTTGG 780
Qy 1500 AGGCCAAATGCCCAATCAGATTTGCTTTATCTCTCAAACTTCCATCTTATGATGACATCT 1559
Db 781 AGGCCAAATGCCCAATCAGATTTGCTTTATCTCTCAAACTTCCATCTTATGATGACATCT 840
Qy 1560 AATTCCTCAATGGAACTGGAGATATCATTTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1619
Db 841 AATTCCTCAATGGAACTGGAGATATCATTTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Qy 1620 TAAAGGTATCAACAGAAACTTGGAAACAGGCTTTATCCCTCTCAAAAGTCCGAGA 1679
Db 901 TAAAGGTATCAACAGAAACTTGGAAACAGGCTTTATCCCTCTCAAAAGTCCGAGA 960
Qy 1680 GAAGATAGAAACAGTCAAG 1698
Db 961 GAAGATAGAAACGGTCAAG 979
```

## RESULT 10

ABK70037

ID ABK70037 standard; cDNA; 979 BP.

XX AC

XX ABK70037;

XX AC

DT 15-JUL-2002 (first entry)

XX AC

DE Antibody production method related cDNA #5.

XX AC

KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.

OS Rattus norvegicus.

XX WO200231140-A1.

XX 18-APR-2002.

XX 05-OCT-2001; 2001WO-JP08804.

XX 06-OCT-2000; 2000JP-0308526.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;

XX Hosaka E, Yamano K, Yamasaki M, Hanai N;

XX WPI; 2002-340182/37.

XX Cells producing antibody compositions including antibody fragments and  
XX fusion proteins with Fc domain of antibody, useful for prevention or  
XX treatment of cancer, immune diseases, circulatory diseases and  
XX infections

XX Example 9; Page 13-14; 314pp; Japanese.

XX This invention relates to novel method for antibody production  
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
XX transferred with a gene encoding an antibody molecule for producing a  
XX composition comprising an antibody molecule with an Fc domain bonded  
XX to the N-glycoside linkage complex sugar chain. The produced antibody  
XX compositions are drugs for prevention or treatment of diseases  
XX accompanying tumour, allergy or inflammation, autoimmune diseases,  
XX circulatory diseases, and viral and bacterial infections. The  
XX antibodies can be stably produced using the method of the invention  
XX with high binding activity and potency thus leading to high safety and  
XX reduced side effects when applied alone or in combination with other  
XX drugs for therapy. The present sequence represents a nucleotide  
XX molecule used in the method of the invention.

SQ Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;

## Query Match

Best Local Similarity 53.9%; Score 931; DB 24; Length 979;

Matches 949; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 720 ACTCATTTGGAATCTCAGAAATGGCGCTATGCTACTGGTGGATGGGAGACTGTTTGG 779  
Db 1 ACTCATTTGGAATCTCAGAAATGGCGCTATGCTACTGGTGGATGGGAGACTGTTTGG 60

Qy 780 ACCTGTAAGTGAGACATGTACAGACATCTGGGCTCTCCACTGGACACTGGTCAGGTGA 839  
Db 61 ACCTGTAAGTGAGACATGTACAGACATCTGGGCTCTCCACTGGACACTGGTCAGGTGA 120

Qy 840 AGTAAATGACAAAAAATTCAGAGTGGTGGAGTCCCATTTGTAGACAGCTCCATCTCTCG 899  
Db 121 AGTAAATGACAAAAAATTCAGAGTGGTGGAGTCCCATTTGTAGACAGCTTCATCTCTCG 180

Qy 900 GCCTCCTTACTTACCAGTGGCTGTTCCAGAGAGCTTTCAGAGACCGACTCTCTAAAGAGTCCA 959  
Db 181 GCCTCCTTACTTACCAGTGGCTGTTCCAGAGAGCTTTCAGAGACCGACTCTCTAAAGAGTCCA 240

Qy 960 TGGTGACCTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1019  
Db 241 TGGTGATCTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300

Qy 1020 TTGGCTGGAAGGAAATAGAGAGAGCCACCAAGAGCTTGGCTTCAAACTATCCAGTTAT 1079  
Db 301 TTGGCTGGAAGGAAATAGAGAGAGCCACCAAGAGCTTGGCTTCAAACTATCCAGTTAT 360

QY 1080 TGGAGTCCATGTCAGACGCACAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGA 1139  
DB |||||  
361 TGGAGTCCATGTCAGACGCACAGACAAAGTGGGAACAGAGCAGCTTCCATCCCATCGA 420  
QY 1140 GGAGTACATGTTACAGCTTGAAGAACATTTTCAGCTTTCGACGCAGCAATGCAAGTGA 1199  
DB |||||  
421 AGAGTACATGTTACAGCTTGAAGAACATTTTCAGCTTTCGACGCAGCAATGCAAGTGA 480  
QY 1200 TAAAAAAGAGTATCTGGCTTACTGATGATCTCTCTTTTAAAGGAGGCAACACAAA 1259  
DB |||||  
481 TAAAAAAGAGTATCTGGCTTACTGATGATCTCTCTTTTAAAGGAGGCAACACAAA 540  
QY 1260 GTACTCCAAATATGAATTTATTTAGTGAATCACTTTCTTTGGTTCAGCTGCACTACAAA 1319  
DB |||||  
541 GTACTCCAAATATGAATTTATTTAGTGAATCACTTTCTTTGGTTCAGCTGCACTACAAA 600  
QY 1320 TCGGTACACAGAAATTCACCTCGGGGTGATCTCTGATATACACTTTCTCTCAGGC 1379  
DB |||||  
601 TCGGTACACAGAAATTCACCTCGGGGTGATCTCTGATATACACTTTCTCTCAGGC 660  
QY 1380 TGACTTTCTAGTGTCTACTTTTTCATCCAGGTCTGTGGGTGCTTATGAATCATGCA 1439  
DB |||||  
661 TGACTTTCTAGTGTCTACTTTTTCATCCAGGTCTGTGGGTGCTTATGAATCATGCA 720  
QY 1440 AACCTGCATCTGATGCTCTCGAACTTCCATTTCTTTGGATGACATCTACTATTTGG 1499  
DB |||||  
721 AACCTGCATCTGATGCTCTCGAACTTCCATTTCTTTGGATGACATCTACTATTTGG 780  
QY 1500 AGGCCAAATGCCCAATCAGATGCTGTTTATCTCTCAAACTTTCAGTACATCTACTATTTGG 1559  
DB |||||  
781 AGGCCAAATGCCCAATCAGATGCTGTTTATCTCTCAAACTTTCAGTACATCTACTATTTGG 840  
QY 1560 AATTCCAATGAACTGAGATATCATTTGGTGTGCTGGAACCACTTGGATGTTTTC 1619  
DB |||||  
841 AATTCCAATGAACTGAGATATCATTTGGTGTGCTGGAACCACTTGGATGTTTTC 900  
QY 1620 TAAAGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGA 1679  
DB |||||  
901 TAAAGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGA 1679  
QY 1680 GAAGATAGAAACGTCAG 1698  
DB |||||  
961 GAAGATAGAAACGTCAG 979

## RESULT 11

AAC63891  
ID AAC63891 standard; cDNA; 979 BP.

XX AAC63891;

DT 09-FEB-2001 (first entry)

XX Chinese hamster FUT8 cDNA.

XX Immunologically functional molecule; immune system; immunomodulation;  
KW glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;  
KW viral infection; bacterial infection; allergy; autoimmune disease;  
KW inflammation; antibody; Chinese hamster; FUT8; ss.

XX Cricetulus griseus.

XX WO200061739-A1.

XX 19-OCT-2000.

XX 07-APR-2000; 2000WO-JP02260.

XX 09-APR-1999; 99JP-0103158.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

PI Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;

PI Imabeppu S, Kanda Y, Yemane N, Anazawa H;  
XX WPI; 2000-665129/64.  
XX Control of activity of antibodies and other immunologically functional  
PT molecules by addition or removal of fucose from sugar chain for  
PT diagnosis and treatment of cancer, allergy and other diseases -  
XX Example 8; Page 72-73; 81pp; Japanese.  
XX The invention relates to a method for controlling the activity of an  
CC immunologically functional molecule (e.g., an antibody) where the  
CC control is effected by the presence or absence of fucose bound to an  
CC N-acetylglucosamine residue at the reducing end of the sugar chain on  
CC the immunologically functional molecule. The invention also relates to  
CC methods for the diagnosis, prevention or treatment of diseases which  
CC involve the modified immunologically functional molecule, and agents  
CC which stimulate the activity of an immunologically functional molecule.  
CC The methods of the invention are used for the diagnosis, treatment and  
CC prevention of a broad range of diseases including cancer, circulatory  
CC disease, viral or bacterial infection, allergy, autoimmune disease and  
CC inflammation. The present sequence represents a Chinese hamster FUT8  
CC cDNA isolated in an exemplification of the invention.

XX Sequence 979 BP; 286 A; 227 C; 218 G; 248 T; 0 other;

Query Match 51.7%; Score 894.2; DB 21; Length 979;

Best Local Similarity 94.6%; Pred. No. 2.2e-261;  
Matches 926; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 720 ACTCATCTTGGAACTTCAGAAATGGCGCTATGCTACTGCTGGATGGGAGACTGTTTAG 779

DB 1 ACTCATCTTGGAACTTCAGAAATGGCGCTATGCTACTGCTGGATGGGAGACTGTTTAG 60

QY 780 ACCTGTAAGTACAGATGTAACAGAGATCGGCTCTCCACTGACACTGGTTCAGGTGA 839

DB 61 ACCTGTAAGTACAGATGTAACAGAGATCGGCTCTCCACTGACACTGGTTCAGGTGA 120

QY 840 AGTAATGACAAAACATTCAGATGGTGGCTCCCACTTCTAGACAGCTCCATCTCG 899

DB 121 AGTAATGACAAAACATTCAGATGGTGGCTCCCACTTCTAGACAGCTCCATCTCG 180

QY 900 GCCTCTTCTTACCTACCTGCTGTTCCAGAAAGACCTTCAGACAGCTTCTTAAGAGTCCA 959

DB 181 TCCTCTTCTTACCTACCTGCTGTTCCAGAAAGACCTTCAGACAGCTTCTTAAGAGTCCA 240

QY 960 TGGTGAACCTGCACTGCTGGGTGTCCTCAGTTGTTCAAACTTCTGATTCGTCACACC 1019

DB 241 TGGTGAACCTGCACTGCTGGGTGTCCTCAGTTGTTCAAACTTCTGATTCGTCACACC 300

QY 1020 TTGGCTGAAAGGAAATAGAAAGCCCAAGAGCTTGGCTTCAAACTTCTGATTCGTCACACC 1079

DB 301 TTGGCTGAAAGGAAATAGAAAGCCCAAGAGCTTGGCTTCAAACTTCTGATTCGTCACACC 360

QY 1080 TGGAGTCCATCTCAGACGCACAGACAAAGTGGGAACAGAGCAGCTTCCACCCCATCGA 1139

DB 361 TGGAGTCCATCTCAGACGCACAGACAAAGTGGGAACAGAGCAGCTTCCATCCCATGGA 420

QY 1140 GGAGTACATGTTACAGCTTGAAGAACATTTTCAGCTTCTCGACGCAGAGTGAAGTGA 1199

DB 421 GGAATACATGTTACAGCTTGAAGAACATTTTCAGCTTCTCGACGCAGAGTGAAGTGA 480

QY 1200 TAAAAAAGAGTATCTGGCTTACTGATGATCTTCTTTTAAAGGAGGCAACACAAA 1259

DB 481 TAAAAAAGAGTATCTGGCTTACTGATGATCTTCTTTTAAAGGAGGCAACACAAA 540

QY 1260 GTACTCCAAATATGAATTTATTTAGTGAATCACTTCTTTGGTTCAGCTGCACTACAAA 1319

DB 541 GTACTCCAAATATGAATTTATTTAGTGAATCACTTCTTTGGTTCAGCTGCACTACAAA 600

QY 1320 TCGGTACACAGAAATTCACCTCGGGGTGATCTCTGATATACACTTTCTCTCAGGC 1379

DB 601 CCGATACACAGAAATTCACCTCGGGGTGATCTCTGATATACACTTTCTCTCAGGC 660

CC antibodies can be stably produced using the method of the invention  
CC with high binding activity and potency thus leading to high safety and  
CC reduced side effects when applied alone or in combination with other  
CC drugs for therapy. The present sequence represents a nucleotide  
CC molecule used in the method of the invention.  
XX  
SQ Sequence 979 BP; 286 A; 227 C; 218 G; 248 T; 0 other;

Query Match 51.7%; Score 894.2; DB 24; Length 979;  
Best Local Similarity 94.6%; Pred. No. 2.2e-261;  
Matches 926; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 720 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGGTGGAGGAGACTGTTTGG 779  
Db 1 ACTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGGAGGATGGGAGACTGTTTGG 60

QY 780 ACCTGTAAGTGAGACATGTACAGACAGATCTGGGCTCTCCACTGGACACTGGTCAAGTGA 839  
Db 61 ACCTGTAAGTGAGACATGTACAGACAGATCTGGGCTCTCCACTGGACACTGGTCAAGTGA 120

QY 840 AGTAAATGACAAAAACATTTCAAGTGTGGAGTCCCATTTGTAGACAGCTCCATCTCG 899  
Db 121 AGTGAAGACAAAAATGTTCAAGTGTGGAGTCCCATTTGTAGACAGCTCCATCTCG 180

QY 900 GCCTCCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACCGACTCTCTAAGAGTCCA 959  
Db 181 TCCTCCTTACTTACCCTTGGCTGTACCAGAGACCTTGCAGATCGACTCTCTGAGAGTCCA 240

QY 960 TGGTGACCTGTCAGTGTGGGTGTCCTCAATTTGTCAAATATCTTGAATTCCTCCACAACC 1019  
Db 241 TGGTGATCTGTCAGTGTGGGTGTCCTCAATTTGTCAAATATCTTGAATTCCTCCACAACC 300

QY 1020 TTGGCTGAAAAAGAAATAGAGAAAGCCACCAAGAGCTTGGCTTCAAATCCAGTTAT 1079  
Db 301 TTGGCTGAAAAAGAAATAGAGAAAGCCACCAAGAGCTTGGCTTCAAATCCAGTTAT 360

QY 1080 TGGAGTCCATGTACAGCGCACAGAAAGTGGGACAGAGACAGCTTCCACCCCATCGA 1139  
Db 361 TGGAGTCCATGTACAGCGCACAGAAAGTGGGACAGAGACAGCTTCCATCCCATCGA 420

QY 1140 GGAGTACATGTTGACACGTTGAAAGAACATTTTCAGCTTCTCCAGCGAGAAATGCAAGTGA 1199  
Db 421 GGAATACATGTTGACACGTTGAAAGAACATTTTCAGCTTCTCCAGCGAGAAATGCAAGTGA 480

QY 1200 TAAAAAAGAGATATATCTGGCTACTGATGCTCTCTTTGTTAAAGAGGCAAGACAAA 1259  
Db 481 TAAAAAAGAGATATATCTGGCTACTGATGCTCTCTTTGTTAAAGAGGCAAGACAAA 540

QY 1260 GTACTCCAATTTATGAATTTATTTAGTGATTAATCTTCTTGGTTCAGTGGACTACAAA 1319  
Db 541 GTACTCCAATTTATGAATTTATTTAGTGATTAATCTTCTTGGTTCAGTGGACTACAAA 600

QY 1320 TCGGTACACAGAAATTCACCTTCGGGGTGTGATCTCTGGATATACATTTCTCTCACAGGC 1379  
Db 601 CCGATACACAGAAATTCACCTTCGGGGTGTGATCTCTGGATATACATTTCTCTCACAGGC 660

QY 1380 TGACTTTCTAGTGTGTTACTTTTTCATCCAGGTCTGCGGGTGTGCTTATGAATATCATGA 1439  
Db 661 TGACTTTCTAGTGTGTTACTTTTTCATCCAGGTCTGCGGGTGTGCTTATGAATATCATGA 720

QY 1440 AACCTGTCATCTGATGCTCTGCGAACTTCCATTTCTTGGATGACATCTACTATTTGG 1499  
Db 721 AACCTGTCATCTGATGCTCTGCGAACTTCCATTTCTTGGATGACATCTACTATTTGG 780

QY 1500 AGGCAAAATGCCCAACATCAGATTGCTGTTTATCTCTCAAAACCTCGAACTGAAGAGGA 1559  
Db 781 AGGCAAAATGCCCAACATCAGATTGCTGTTTATCTCTCAAAACCTCGAACTGAAGAGGA 840

QY 1560 AATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCAATCGGATGTTTTC 1619  
Db 841 AATCCCATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCAATCGGATGTTTTC 900

QY 1620 TAAAGGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGA 1679  
Db 901 TAAAGGTATCAACAGAAACTTGGAAACAGGCTTATATCCCTCTCAAAAGTCCGAGA 960

QY 1680 GAAGATAGAAACAGTCAAG 1698  
Db 961 GAAGATAGAAACGGTCAAG 979

RESULT 12  
ABK70036  
ID ABK70036 standard; cDNA; 979 BP.  
XX  
AC ABK70036;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Antibody production method related cDNA #4.  
XX  
KW Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection; primer; ss.  
XX  
OS Cricetulus griseus.  
XX  
PN WO200231140-A1.  
XX  
PD 18-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-JP08804.  
XX  
PR 06-OCT-2000; 2000JP-0308526.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI Hosaka E, Yanano K, Yamasaki M, Hanai N;  
XX  
DR WPI; 2002-340182/37.  
XX  
PT Cells producing antibody compositions including antibody fragments and  
PT fusion proteins with Fc domain of antibody, useful for prevention or  
PT treatment of cancer, immune diseases, circulatory diseases and  
PT infections -  
XX  
PS Example 9; Page 12-13; 314pp; Japanese.  
XX  
CC This invention relates to novel method for antibody production  
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell  
CC transfected with a gene encoding an antibody molecule for producing a  
CC composition comprising an antibody molecule with an Fc domain bonded  
CC to the N-glycoside linkage complex sugar chain. The produced antibody  
CC compositions are drugs for prevention or treatment of diseases  
CC accompanying tumour, allergy or inflammation, autoimmune diseases,  
CC circulatory diseases, and viral and bacterial infections. The

```
Db 901 TAAAGGTGTCACAGAAAACTAGGAAAAACAGCGCTGTACCTTCTCTACAAAGTCCGAGA 960
Qy 1680 GAAGATAGAAACAGTCAAG 1698
Db 961 GAAGATAGAAACGGTCAAG 979

RESULT 13
AAF87952
ID AAF87952 standard; DNA; 1017 BP.
AC AAF87952;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 237-575 encoding DNA SEQ ID NO:1.
XX
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;
KW ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1..1017
FT CDS
FT /tag= a
FT /partial
FT /product= "alpha 1-6 fucosyltransferase 237-575"
XX
PN JP2001011097-A.
XX
PD 16-JAN-2001.
XX
PF 29-JUN-1999; 99JP-0183569.
XX
PR 29-JUN-1999; 99JP-0183569.
XX
PA (FJRE ) FUJIREBIO KK.
XX
WPI; 2001-275926/29.
DR P-PSDB; AAF875061.
XX
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunoassay -
XX
PS Example 1; Page 6-7; 11pp; Japanese.
XX
CC The present invention describes an anti-human alpha 1-6
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose
CC transferase) antibody. Also described are: (1) a hybridoma producing
CC the above monoclonal antibody; (2) an immunoassay for detecting human
CC alpha 1-6 fucT by using the above antibody or its antibody fragment;
CC and (3) a reagent used for the above immunoassay. The anti-human
CC alpha 1-6 fucT antibody can be used for immunoassay. The present
CC sequence encodes the human alpha 1-6 fucT protein of residues 237 to
CC 575, which is used in an example from the present invention.
XX
SQ Sequence 1017 BP; 296 A; 231 C; 228 G; 262 T; 0 other;

Query Match 50.7%; Score 876.2; DB 22; Length 1017;
Best Local Similarity 91.3%; Pred. No. 6.8e-256;
Matches 929; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 709 ACCCAGCGAACACTCATCTTGAATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAG 768
Db 1 ACCCAGCGAACACTCATCTTGAATCTCAGAAATGGCGCTATGCTACTGGTGGATGGAG 60
Qy 769 ACTGTGTTAGACCTGTAGTACAGACATGTACAGACAGATCTGGGCTCTCCACTGGACAC 828
Db 61 ACTGTATTTAGGCGCTGTAGTACAGACATGTACAGACAGATCTGGGCTCTCCACTGGACAC 120
Qy 829 TGGTCAGGTGAAGTAAATGACAAAAACATTCAAGTGGTTCGAGCTCCCGCTTGTAGACAGC 888
```

```
Db 121 TGGTCAGGTGAAGTGAAGCAGAAAAATGTTCAAGTGGTGCAGCTTCCCTATTGTAGACAGT 180
Qy 889 CTCATCTCGGCGCTCTTACTTACCACTGGCTGTTCAGAGAGACCTTGCAGACCGACTC 948
Db 181 CTTTCATCCCGTCCCTCATATTTTACCCCTGGCTGTACCAGAGAGACCTCGAGATGCATT 240
Qy 949 CTAAGAGTCCATGGTGACCCCTGCAGTGTGGTGGTGTCCCGAGTTTGTCAAATPACTTGATT 1008
Db 241 GTACAGTGCATGGTGACCCCTGCAGTGTGGTGTCTCAGTTTGTCAAATPACTTGATC 300
Qy 1009 CGTCCAAACCTTGGCTGGAAAAAGAAATAGAGAAGCCACCAAGAGCTTGGCTTCAAA 1068
Db 301 CGCCACACAGCCTTGGCTAGAAAAAGAAATAGAGAAGCCACCAAGAGCTTGGCTTCAAA 360
Qy 1069 CATCCAGTTATTGGAGTCCATGTCCAGCGCACAGACAAAGTGGGAACAGAGCACCCCTTC 1128
Db 361 CATCCAGTTATTGGAGTCCATGTCCAGCGCACAGACAAAGTGGGAACAGAGCTGCCCTTC 420
Qy 1129 CACCCATCGAGGAGTACATGGTACACGTTGAAGAAACATTTTTCAGCTTCTCCGACGACGA 1188
Db 421 CATCCATTTGAAGAGTACATGGTGCATGTTGAAGAACATTTTCAGCTTCTTCGACGACGA 480
Qy 1189 ATGCAAGTGGATAAAAAAGAGATATCTGGCTACTGATGATCTCTTTGTAAAGGAG 1248
Db 481 ATGCAAGTGGACAAAAAAAGAGTGTATTTGGCCACAGATGACCCCTCTCTTATTAAGGAG 540
Qy 1249 GCAAAGACAAAGTACTCCAATTATGAATTTATTAGTGAATACTCTATTTCTTGGTCAGCT 1308
Db 541 GCAAAAACAAAGTACCCCAATTTATGAATTTATTAGTGAATACTCTATTTCTTGGTCAGCT 600
Qy 1309 GGACTACACAATCGGTACACAGAAAAATTCATTTCCGGGTGTGATCCTCGATATACACTTT 1368
Db 601 GGACTGACAAATCGATACACAGAAAAATTCATTTCCGGGTGTGATCCTCGATATACACTTT 660
Qy 1369 CTCTCACAGGCTGACTTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTTCCTTAT 1428
Db 661 CTCTCTCAGGCAGACTTCTAGTGTGATCTTTTTCATCCAGGTCTGTGGGTTCCTTAT 720
Qy 1429 GAAATCATGCAACCCCTGCATCCTGATGCCCTCTCGAACTTCCATCTTCTTGATGACATC 1488
Db 721 GAAATTCATGCAACCACTACATCCTGATGCCCTCTCGAACTTCCATCTTCTTGATGACATC 780
Qy 1489 TACTATTTTGGAGCGCAAAATGCCACAATCAGATTTGCTGTTTATCTCTCAAAACCTCGA 1548
Db 781 TACTATTTTGGAGCGCAAAATGCCACAATCAGATTTGCTGTTTATCTCTCAAAACCTCGA 840
Qy 1549 ACTGAAGAGGAAATTCCAATGGAACCTGGAGATATCATTTGGTGGCTGGAAACCATTTGG 1608
Db 841 ACTGCAGATGAAATTCCTCATGGAACCTGGAGATATCATTTGGTGGCTGGAAATCATTTGG 900
Qy 1609 GATGGTTATTTCTAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTTAC 1668
Db 901 GATGGCTATTTCTAAAGGTATCAACAGGAAATTTGGGAAGGAGCGGCTATATCCCTCTTAC 960
Qy 1669 AAGTCCGAGAGAGATAGAACAGTCAAGTATCCACATATCCCATATCTCTGAGCTGAAAAA 1725
Db 961 AAGTTCGAGAGAGATAGAAAAACGTCAGTACCCCAATATCTCTGAGGCTGAGAAA 1017

RESULT 14
AAF87953
ID AAF87953 standard; DNA; 699 BP.
XX
AC AAF87953;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 343-575 encoding DNA SEQ ID NO:2.
XX
KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay;
KW ds.
```

```

XX OS Homo sapiens.
XX FH Key
XX FT CDS
XX FT Location/Qualifiers
XX FT 1..699
XX FT /*tag= a
XX FT /partial
XX FT /product= "alpha 1-6 fucosyltransferase 343-575"
XX PN JP2001011097-A.
XX PD 16-JAN-2001.
XX PF 29-JUN-1999; 99JP-0183569.
XX PR 29-JUN-1999; 99JP-0183569.
XX PA (FJRE ) FUJIREBIO KK.
XX DR WPI; 2001-275926/29.
XX DR P-PSDB; AAB75062.
XX PT Novel anti-human alpha1-6 fucose transferase antibody useful for
XX PT immunosassay -
XX PS Example 1; Page 7-8; 11pp; Japanese.
XX CC The present invention describes an anti-human alpha 1-6
XX CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
XX CC transferase) antibody. Also described are: (1) a hybridoma producing
XX CC the above monoclonal antibody; (2) an immunoassay for detecting human
XX CC alpha 1-6 fuct by using the above antibody or its antibody fragment;
XX CC and (3) a reagent used for the above immunoassay. The anti-human
XX CC alpha 1-6 fuct antibody can be used for immunoassay. The present
XX CC sequence encodes the human alpha 1-6 fuct protein of residues 343 to
XX CC 575, which is used in an example from the present invention.
XX SQ Sequence 699 BP; 220 A; 152 C; 147 G; 180 T; 0 other;

Query Match 35.0%; Score 604.6; DB 22; Length 699;
Best Local Similarity 91.6%; Pred. No. 2.9e-173;
Matches 640; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1027 GAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTATTGGAGTC 1086
DB 1 GAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAACATCCAGTATTGGAGTC 60

QY 1087 CATGTACAGCCACAGACAAAGTGGGACAGAGCAGCTTCCACCCATCGAGAGTAC 1146
DB 61 CATGTACAGCCACAGACAAAGTGGGACAGAGCTTCCATCCCATTTGAAGAGTAC 120

QY 1147 ATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGATAAAAA 1206
DB 121 ATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGACAAAAA 180

QY 1207 AGAGTATATCTGGCTACTGATCCTACTTTTGTAAAGGCGGACAAAGTACTCC 1266
DB 181 AGAGTATATCTGGCTACTGATCCTACTTTTGTAAAGGCGGACAAAGTACTCC 240

QY 1267 AATTATGAATTTATTAGTGATTAACCTATTCTTGGTCACTGGAGTACACAAATCGGTAC 1326
DB 241 AATTATGAATTTATTAGTGATTAACCTATTCTTGGTCACTGGAGTACACAAATCGGTAC 300

QY 1327 ACAGAAAATTCATTTGGGGTGTGATCTCTGGATATACATTTCTCTCAGAGCTCACTTT 1386
DB 301 ACAGAAAATTCATTTGGGGTGTGATCTCTGGATATACATTTCTCTCAGAGCTCACTTT 360

QY 1387 CTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTGCTTATGAAATCATGCAAAACCTGTG 1446
DB 361 CTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTGCTTATGAAATCATGCAAAACCTA 420

QY 1447 CATCTGATGCTCTCGGAGTCCATCTTTGGATGACATCTACTATTTTGGAGGCCAA 1506

```

```

DB 421 CATCTGATGCTCTGCAGAACTTCCATTTCTTTAGATGACATCTACTATTTTGGGGCCAG 480
QY 1507 AATGCCACAAATCAGATTGCTGTTTATCTCTCACAACCTCGAACTGAAGAGAAATTCCA 1566
DB 481 AATGCCACAAATCAGATTGCTGTTTATCTCTCACAACCTCGAACTGAAGAGAAATTCCA 540
QY 1567 ATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGGATGTTTCTTAAAGGT 1626
DB 541 ATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGGATGTTTCTTAAAGGT 600
QY 1627 ATCAACAGAAAACCTTGGAAAACAGGGTTATATATCCCTCTACAAAGTCCGAGAGAAGATA 1686
DB 601 GTCAACAGAAAATTTGGGAGGACGGCTATATCTCTCTACAAAGTTCGAGAGAAGATA 660
QY 1687 GAAACAGTCAAGTATCCACATATCTCTGAAAGCTGAAANA 1725
DB 661 GAAACGGTCAAGTATCCACATATCTCTGAGGCTGAGAAA 699

RESULT 15
ABL04601
ID ABL04601 standard; cDNA; 2761 BP.
XX AC ABL04601;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 8285.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li FWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB60498.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 8285; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2761 BP; 657 A; 731 C; 755 G; 618 T; 0 other;

Query Match 21.9%; Score 378.8; DB 23; Length 2761;
Best Local Similarity 57.2%; Pred. No. 3.9e-104;
Matches 711; Conservative 0; Mismatches 522; Indels 9; Gaps 1;

```



Qy	473	ACCATGAAGGCTCTATCATGACAGATCTATCTACTCCTCAGTCTCAAAACAGATGTGGAGCAGGGG	532
Db	1083	AGCACAAGCGCTCCCTCTGAGCGGATATGAGCGGATGCGTCACTCCGATGGCTTACGAGG	1142
Qy	533	ATTGGCGTGAANAAGAGGCCAABAGATCTGACAGAGCTGGTCCAGCGGAGAAATCAATATC	592
Db	1143	CCTGGCGACACAAGGAGCGGGAATTTAGCGAATTTGGTACAGCGCGCTGCACCATC	1202
Qy	593	TCCAGAAATCCTTAAGGACTGACAGCAAAAGCCAGGAAGCTGGTGTGTAAATCAATCAATAAAGGCT	652
Db	1203	TACAGAATCCAGTGAATTCGCAGAAATGCTCGAAGCTGGTCTGTAAAGCTCAACAAGGCT	1262
Qy	653	GTGGCTATGGTTGTCAACTCATCAGCTGGTCTACTGTTTTCATGATTTGCTTTATGGCAACC	712
Db	1263	GTGGCTACGGGTGCAACTGACCATGTGGTGTATTGCTTTATAGTGGCCCTACGCCACCG	1322
Qy	713	AGCGAACACTCATCTCTGGAATCTCAGAAATTTGGGCTATGCTACTGGTGGATGGGAGACTG	772
Db	1323	AGCGGACGCTCACTTAAATCCCGCGGATGGCGGTATCAAGGGCGGCTGGGAGGAGG	1382
Qy	773	TGTTTAGACCTGTAAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGGGACACTGGT	832
Db	1383	TGTTTCCAAACGGTGTCCAAACAGCTGCCATGATGGGGCACCGCAACACATACAAATTTGGC	1442
Qy	833	CAGGTGAAGTAAATGACAAAAAACTTCAAGTGTGAGCTCCCGATTTGTATAGACAGCCTCC	892
Db	1443	CGGGCA-----AGCCGAACACCCAGGTGTGGTGTGGCCCATCATTTGACTCGCTGA	1493
Qy	893	ATCCTCGGCTCCCTTACTTACCACCTGCTGTTCACAGAAGACCTTTCAGACCGGACTCCTAA	952
Db	1494	TGCCGAGACCGCGGTACTCGCGCTCGCGTTCCGAGGATCTGGCGCCGAGACTTAAAGC	1553
Qy	953	GAGTCCATGTGTGACCCCTGCAGTGTGGTGTGCCAGTTTGTCAAATACTTGAATTCGTC	1012
Db	1554	GTCTGCATGAGACCCCATCGTTTGTGGTGGTGCAGTTTCTTAAGTATTTGCTCGCAC	1613
Qy	1013	CACAACTTGGCTGGGAAAAGGAATAGAGAGCCACAGAGCTTGGCTTCAAACATC	1072
Db	1614	CGCAACCAACGACCGGGATTTTCTTACTCTGGCATGCGCAATTTGGGTGGGAGCGTC	1673
Qy	1073	CAGTTATTGGAGTCCATGTACAGCGCACAGACAAAGTGGGAAACAGAGACGCTTCCACC	1132
Db	1674	CCATTGTTGGCTTCATGTCGTGCGACGACAAAGTGGGCAACAGAGCGGCTGCCACA	1733
Qy	1133	CCATCGAGGAGTACATGGTACAGTTGAAGAACATTTTTCAGTCTTCGCAACGAGAAATGC	1192
Db	1734	GTGTGGAGGAGTATATGACCTATGTCGAGGACTACTACCGCACGCTGGAGGTAAACGGCA	1793
Qy	1193	AAGTGGATAAAAAAGAGTATATCTGGCTACTGATGATCTACTTTGTTAAAGGAGCA	1252
Db	1794	GCACTGTGGCTCGCCGATTTCTCGCATCGGATGATGCCAGGTTATTGAGGAGGCGC	1853
Qy	1253	AGACAAAGTACTCCAAATATGAATTTATAGTGTATAACTCTATTCTTGTGTGAGCTGGAC	1312
Db	1854	GCCGAAGTACCCGCGTACCAAGTCAATGGTGATCGGAGGTGGCGCGTATGGCGTCCG	1913
Qy	1313	TACAAATCGGTACACAGAAAATTCATCTGGGGGTGTGATCTGTGGATATACACTTCTCT	1372
Db	1914	TATCCACACGATACACGGACACCGCCCTGAAACGGGATCATTTCTGGATATCCACTGCTTT	1973
Qy	1373	CACAGGCTGACTTCTAGTGTGTACTTTTTCATCCAGGTCGTGCGGGTGTCTATGAAA	1432
Db	1974	CCATGTCGATCATCTGGTGTGACCTTCTCGTCGAGGTTGTGCGGTGGCCTACGAGA	2033
Qy	1433	TCATGCAAAACCTGCATCCTGATGCCCTCTGCAACTTCCATTTCTTTGGATGACATCTACT	1492
Db	2034	TAATGCAGACGATATCCGGATGCGGGATCGGTTCAAGTCGCTGACGACATATACT	2093
Qy	1493	ATTTTGGAGGCCAAAATGCCCAAAATCAGATTTGCTGTTTATCTCTCAAAACCTGGAACCTG	1552
Db	2094	ACTACGGTGGCAGAAATCGCAAAATCGCGCGTCTGTTATCGGCCCAACAGCGCGCACGC	2153

Search completed: February 2, 2004, 12:33:55  
Job time : 475.615 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 12:05:46 ; Search time 104.069 Seconds  
(without alignments)  
7328.922 Million cell updates/sec

Title: US-09-971-773-2

Perfect score: 1728

Sequence: 1 atcgggcgactgactgttc.....atctgaagctgaaaaatag 1728

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496.6	86.6	2100	3	US-08-913-805A-9 Sequence 9, Appli
2	1496.6	86.6	2100	3	US-09-442-629-9 Sequence 9, Appli
3	1469.4	85.0	1728	3	US-08-913-805A-1 Sequence 1, Appli
4	1469.4	85.0	1728	3	US-09-442-629-1 Sequence 1, Appli
5	74	4.3	7218	1	US-08-232-463-14 Sequence 14, Appli
6	38.2	2.2	921	4	US-09-328-352-2691 Sequence 2691, Ap
7	38	2.2	1230025	4	US-09-198-452A-1 Sequence 1, Appli
8	37.6	2.2	2116	3	US-09-000-041A-1 Sequence 1, Appli
9	37.2	2.2	1596	4	US-09-134-001C-2083 Sequence 54, Appli
10	36.8	2.1	1750	2	US-08-933-750C-54 Sequence 54, Appli
11	36.8	2.1	1750	3	US-09-234-613-54 Sequence 9, Appli
12	36	2.1	4517	3	US-09-140-804-9 Sequence 9, Appli
13	36	2.1	4517	4	US-09-686-838B-9 Sequence 9, Appli
14	36	2.1	4517	4	US-09-776-976-5 Sequence 5, Appli
15	36	2.1	4517	4	US-09-909-547-5 Sequence 5, Appli
16	35.8	2.1	1664976	4	US-08-916-421B-1 Sequence 1, Appli
17	35.2	2.0	978	4	US-09-134-001C-110 Sequence 110, Appli
18	35	2.0	313	4	US-09-313-294A-4658 Sequence 4658, Ap
19	34.6	2.0	1664976	4	US-08-916-421B-1 Sequence 1, Appli
20	34.2	2.0	602	4	US-09-538-709-21 Sequence 21, Appli
21	34.2	2.0	1268	4	US-09-538-709-19 Sequence 19, Appli
22	34	2.0	1591	2	US-08-646-981-14 Sequence 14, Appli
23	34	2.0	2915	4	US-09-336-115C-5 Sequence 5, Appli
24	33.8	2.0	805	3	US-08-961-083-139 Sequence 139, Appli
25	33.8	2.0	805	4	US-09-536-784-139 Sequence 139, Appli
26	33.8	2.0	969	4	US-09-252-991A-12688 Sequence 12688, A
27	33.8	2.0	1041	4	US-09-252-991A-12831 Sequence 12831, A

28	33.8	2.0	1629	4	US-09-252-991A-13111 Sequence 13111, A
29	33.8	2.0	12127	4	US-08-961-527-148 Sequence 148, Appli
30	33.4	1.9	2628	1	US-08-143-219-1 Sequence 1, Appli
31	33.2	1.9	3356	1	US-09-379-523-4 Sequence 4, Appli
32	33	1.9	1383	4	US-09-134-001C-2485 Sequence 2485, Ap
33	32.8	1.9	241	4	US-09-397-787-70 Sequence 70, Appli
34	32.8	1.9	1425	4	US-09-592-998C-6 Sequence 6, Appli
35	32.8	1.9	1590	4	US-08-887-534A-73 Sequence 73, Appli
36	32.8	1.9	1590	4	US-09-527-431-73 Sequence 73, Appli
37	32.8	1.9	1656	3	US-08-699-103B-7 Sequence 7, Appli
38	32.8	1.9	1656	4	US-09-229-059-7 Sequence 7, Appli
39	32.8	1.9	1656	4	US-09-628-133-7 Sequence 7, Appli
40	32.8	1.9	2699	4	US-09-336-115C-3 Sequence 3, Appli
41	32.8	1.9	5361	3	US-08-973-462-2 Sequence 2, Appli
42	32.8	1.9	6152	3	US-08-973-462-1 Sequence 1, Appli
43	32.8	1.9	36651	4	US-09-738-894A-3 Sequence 3, Appli
44	32.8	1.9	36651	4	US-09-964-469-3 Sequence 3, Appli
45	32.8	1.9	168575	4	US-09-426-290-1 Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-08-913-805A-9  
; Sequence 9, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 5.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2100 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-913-805A-9

Query Match 86.6%; Score 1496.6; DB 3; Length 2100;  
Best Local Similarity 91.7%; Pred. No. 0; Mismatches 144; Indels 0; Gaps 0;  
Matches 1593; Conservative 0;

QY 1 ATGCGGCGATGAGTGGTTCCTGGGCTTGGATATATGCTCATTTCTTTTTCCTGGGGGACC 60  
DB 198 ATGCGGCGATGAGTGGTTCCTGGGCTTGGATATATGCTCATTTCTTTTTCCTGGGGGACC 257

QY 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCAGATAATATGACCACTTCACCTCCAGC 120  
DB 258 TTGCTGTTTTATATAGTGGTCAATTTGGTTCAGATAATATGACCACTTCACCTCCAGC 317

QY 121 AGAAGACTCTCCAAGATTTCTTCAAGCTTGAACCTTAAACACACAAAATGAAGACTTG 180  
DB 318 CGAAGAACTGTCCAAGATTTCTGCAAGCTTGAACCTTAAACACACAAAATGAAGACTTG 377

QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCATTTGACGAGGACAGCTACA 240  
DB 378 AGGCGAATGGCGAATCTCTCCGATACCAAGAGGCCCATTTGATCAGGGGCCAGCTATA 437

QY 241 GGAAGAGTCCGCTGTTTGAAGAAACAGCTTGTAAAGGCCAAAGAACAGATTTGAAAATTTAC 300  
DB 438 GGAAGAGTACGCGTTTTTGAAGAGCAGCTTGTAAAGGCCAAAGAACAGATTTGAAAATTTAC 497

QY 301 AAGAACAGCTAGAAATGGTCTGGGAGGATCATGAATCTTAAAGAGGAGGATTTGAA 360  
DB 498 AAGAACAGCAGAGAAATGGTCTGGGAGGATCATGAATCTTAAAGAGGAGGATTTGAA 557

QY 361 AATGAGCTTAAAGAGCTCTGGTTTTTTCTTCAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
DB 558 AATGAGCTTAAAGAGCTCTGGTTTTTTCTTCAAGAGTGAATTTGAAGAACTTTA 617

QY 421 GAAGAAATGAATCTCAAAGACATGACAGATGAATTTCTTTGGATTTTAGGACACCATGAA 480  
DB 618 GAAGAAATGAATCTCAAAGACATGACAGATGAATTTCTTTGGATTTTAGGACATCATGAA 677

QY 481 AGGTCTATCATCAGAGTCTATCTACCTCAGTCAACAGATGGAGGAGGATTTGGCGT 540  
DB 678 AGGTCTATATACAGGATCTATCTACCTCAGTCAACAGATGGAGGAGGATTTGGCGG 737

QY 541 GAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGAGAAATTAACATATCTCCAGAAAT 600  
DB 738 GAAAGAGGCGCAAGATCTGACAGAACTGGTTTACGCGAGAAATTAACATATCTCCAGAAAT 797

QY 601 CTTAAGGACTGACGAAAGCCAGGAGCTGGTGTGTAAATCAATCAATTAAGGCTGTGGCTAT 660  
DB 798 CTTAAGGACTGACGAAAGCCAGGAGCTGGTGTGTAAATCAATCAATTAAGGCTGTGGCTAT 857

QY 661 GTTGTCTCACTCATCATGCTGCTTCTGTTTCATGATTTGTCACCCAGCGAACA 720  
DB 858 GGTGTCTCACTCATCATGCTGCTTCTGTTTCATGATTTGTCACCCAGCGAACA 917

QY 721 CTCATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGTGGATGGAGAGCTGTGTTAGA 780  
DB 918 CTCATCTTGGAAATCTCAGAAATTTGGCGCTATGCTACTGTGGATGGAGAGCTGTGTTAGA 977

QY 781 CTTGTAAGTGAACATGTACAGACAGATCTGGCGCTCTCCATCTGACACTGGGTGAGTGA 840  
DB 978 CTTGTAAGTGAACATGTACAGACAGATCTGGCGCTCTCCATCTGACACTGGGTGAGTGA 1037

QY 841 GTAAATGACAAAACATTTCAAGTGTGCGAGCTCCCATTTGTAGACAGCTCCATCTCGG 900  
DB 1038 GTAAATGACAAAACATTTCAAGTGTGCGAGCTCCCATTTGTAGACAGCTCCATCTCGG 1097

QY 901 CCTCTCTACTTACCACTGGCTCTTCCAGAAAGACCTTTGACAGCCGACTCTTAAGAGTCCAT 960  
DB 1098 CCTCTCTACTTACCACTGGCTCTTCCAGAAAGACCTTTGACAGCTGCTGACAGTGCAT 1157

QY 961 GGTGACCCCTGACGTGGTGGGTGCCCGAGTTTGTGCAAAATACTTGAATCTGTCACCAACCT 1020

DB 1158 GGTGACCCCTGACGTGGTGGTCTCAGTTTGTCAAAATCTTGAATCCGCCACAGCT 1217

QY 1021 TGGCTGGAAGAAAGAAATAGAGAACCCACCAAGAGCTTTGGCTTTCAACATCTCAGTTAT 1080

DB 1218 TGGCTAGAAAGAAATAGAGAACCCACCAAGAGCTTTGGCTTTCAACATCTCAGTTAT 1277

QY 1081 GGAGTCCATGTACAGCGACAGACAAAGTGGGAACAGAGAGGCTTCCACCCCATCGAG 1140

DB 1278 GGAGTCCATGTACAGCGACAGACAAAGTGGGAACAGAGAGGCTTCCATCTCATTGAA 1337

QY 1141 GAGTACATGGTACAGCTTGAAGACATTTTCAGCTTCTCGCACGACAGAAATCAAGTGCAT 1200

DB 1338 GAGTACATGGTGCATGTTGAAGACATTTTCAGCTTCTTCGCCGACAGAAATCAAGTGCAC 1397

QY 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTTACTTTTAAAGAGGAGGCAAAAG 1260

DB 1398 AAAAAAGAGTATTTGGCCACAGATGACCTTCTTTTAAAGAGGAGGCAAAAG 1457

QY 1261 TACTCCAATTAAGAAATTTATAGTGAATCTCTATTTCTTGGTCAGCTGGACTACACAAT 1320

DB 1458 TACCCCAATTAAGAAATTTATAGTGAATCTCTATTTCTTGGTCAGCTGGACTGACCAAT 1517

QY 1321 CGGTACACAGAAATTCACCTTCGGGCTGATCTGATATACACTTTCTCTCAGGCT 1380

DB 1518 CGATACACAGAAATTCACCTTCGGGAGTGAATCTGATATACACTTTCTCTCAGGCA 1577

QY 1381 GACTTCTAGTGTACTTTTCAATCCAGGCTCTGTCGGGTTGCTTATGAATCATGCA 1440

DB 1578 GACTTCTAGTGTACTTTTCAATCCAGGCTCTGTCGGGTTGCTTATGAATCATGCA 1637

QY 1441 ACCGTGATCTGATGCTCTGCGAACTTCCATTTCTTTGGATGACATCTACTATTTGGG 1500

DB 1638 ACCTGATCTGATGCTCTGCGAACTTCCATTTCTTTAGATGACATCTACTATTTGGG 1697

QY 1501 GGCCAAAATGCCCAATCAGATTTGCTTTATCTCACAACCTCGAACTGAGAGGAA 1560

DB 1698 GGCCAGAAATGCCCAATCAGATTTGCTTTATCTCACAACCTCGAACTGAGAGGAA 1757

QY 1561 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGATGGTATTCT 1620

DB 1758 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGATGGTATTCT 1817

QY 1621 AAAGGTATCAACAGAAACTTTGGAAACAGGCTTTATATCCCTCTCAAAAGTCCGAGAG 1680

DB 1818 AAAGGTATCAACAGAAACTTTGGAAAGGAGCGGCTTATATCCCTCTCAAAAGTCCGAGAG 1877

QY 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCAGCTGAAATA 1727

DB 1878 AAGATAGAAACAGTCAAGTATCCCAATATCTCAGCTGAAATA 1924

## RESULT 2

US-09-442-629-9  
; Sequence 9, Application US/09442629  
; Patent No. 6291219

## GENERAL INFORMATION:

APPLICANT: TANIGUCHI, Naoyuki  
; UOZUMI, Naofumi  
; SHIBA, Tetsuo

YANAGIDANI, Shusaku

TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

STATE: DC

COUNTRY: US

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-NOV-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2100 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:

US-09-442-629-9

Query Match 86.6%; Score 1496.6; DB 3; Length 2100;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 1583; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 1 ATGGGGCATGGACTGGTCTCTGGGTTGGATTATGCTCATCTCTTTTGGCTGGGGACC 60  
DB |||||  
QY 198 ATGGGGCATGGACTGGTCTCTGGGTTGGATTATGCTCATCTCTTTTGGCTGGGGACC 257  
DB |||||

QY 61 TTGTTATTTTATAGTGGTTCATTGGTTCGAGATATGACCACTTGATCACTCCAGC 120  
DB |||||

QY 258 TTGCTGTTTATAGTGGTTCATTGGTTCGAGATATGACCACTTGATCACTCTAGC 317  
DB |||||

QY 121 AGAGAACTCTCAAGATTCTTGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
DB |||||

QY 318 CGAGAACTGTCCAAGATTCTGGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 377  
DB |||||

QY 181 AGGGAATGGCTGAGTCTCTCGAATACCAAGGCCCCATTTGACAGGGGACAGCTACA 240  
DB |||||

QY 378 AGGGGAATGGGCGAATCTCTCGGATACCAAGGCCCCATTTGATCAGGGGCGCAGCTATA 437  
DB |||||

QY 241 GGAAGATCCGTTGTTAGAGAACAGCTTGAAGGCCAAAGACAGATTGAAATTTAC 300  
DB |||||

QY 438 GGAAGATCCGTTTGAAGAACAGCTTGAAGGCCAAAGACAGATTGAAATTTAC 497  
DB |||||

QY 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGATTGAA 360  
DB |||||

QY 498 AAGAAACAGACCAAGAAATGGTCTGGGGAAGGATCATGAAATCTTGAAGAGGAGATTGAA 557  
DB |||||

QY 361 AATGAGCTAAGAGCTCTGTTTCTTCTCAAGCGAATCAAGAAATTAAGCATTTA 420  
DB |||||

QY 558 AATGAGCTAAGAGCTCTGTTTCTTCTCAAGCGAATCAAGAAATTAAGCATTTA 617  
DB |||||

QY 421 GAAGAAATGAATCCCAAGACATGCAGATGAATCTTTTGGATTAGGACCATGAA 480  
DB |||||

QY 618 GAAAGAAATGAATCCCAAGACATGCAGATGAATCTTTTGGATTAGGACCATGAA 677  
DB |||||

QY 481 AGGTCTATCATGACAGATCTATCTACCTCAGTCAAAACAGATGGAGAGGGGATTGGCGT 540  
DB |||||

DB 678 AGGTCTATTAATGACGGATCTATATACCTCAGTCAGACAGATGGAGCAGGTGATTGGCGG 737  
QY |||||

DB 541 GAAAAGAGGCCAAAGATCTGACAGAGCTGGTCAGCGGAGAAATAACATATCTCCAGAA 600  
DB |||||

DB 738 GAAAAGAGGCCAAAGATCTGACAGAACTGGTTCAGCGGAGAAATAACATATCTTCAGAA 797  
DB |||||

QY 601 CCTAAGGACTGCAGCAAAAGCCAGGAAGCTGGTGTAAACATCAATAAAGGCTGTGGCTAT 660  
DB |||||

DB 798 CCAAGGACTGCAGCAAAAGCCAAAGCTGGTGTAAATATCAACAAAGGCTGTGGCTAT 857  
DB |||||

QY 661 GGTGTCACTCCATCAGCTGCTGCTACTGTTTCAATGATGCTTATGGCACCAGGGAACA 720  
DB |||||

DB 858 GGTGTGAGCTCCATCATGCTGCTACTGCTTCAATGATGCTATGGCACCAGGGAACA 917  
DB |||||

QY 721 CTCACTTTGGAATCTCAGAAATGGCGCTATGCTACTGCTGATGGAGAGCTGTGTTAGA 780  
DB |||||

DB 918 CTCACTTTGGAATCTCAGAAATGGCGCTATGCTACTGCTGATGGAGAGCTGTATTAGG 977  
DB |||||

QY 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCTCTCCACTGGACACTGGTCAGGTGAA 840  
DB |||||

DB 978 CCTGTAAGTGAGACATGTACAGACAGATCTGGCATCTCCACTGGACACTGGTCAGGTGAA 1037  
DB |||||

QY 841 GTAAATGACAAAAACATTTCAAGTGGTGGCTCCCAATTTGAGACAGCTCCATCTCGG 900  
DB |||||

DB 1038 GTGAAGGACAAAAATGTTCAAGTGGTGGCTCCCAATTTGAGACAGCTCTTCATCCCGT 1097  
DB |||||

QY 901 CCTCTTACTTACCACATGGCTGTTCCAGAAAGCTTCAGACAGCTTCCTCAAGAGTCCAT 960  
DB |||||

DB 1098 CCTCATATTTTACCCCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCAT 1157  
DB |||||

QY 961 GGTGACCTCTGAGTGTGGTGGTCTCCAGTTTGTCAAATACTTTGATTCGTCCCAACCT 1020  
DB |||||

DB 1158 GGTGACCTCTGAGTGTGGTGGTCTCCAGTTTGTCAAATACTTTGATTCGTCCCAACCT 1217  
DB |||||

QY 1021 TGGCTGGAAGGAAAAAGAGAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTAT 1080  
DB |||||

DB 1218 TGGCTGGAAGGAAAAAGAGAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTAT 1277  
DB |||||

QY 1081 GGAGTCCATGTGACAGCAGACAGCAAAAGTGGAAAGAGAGCTTCCACCCCATCGAG 1140  
DB |||||

DB 1278 GGAGTCCATGTGACAGCAGACAGCAAAAGTGGAAAGAGAGCTTCCATCCCATGAA 1337  
DB |||||

QY 1141 GAGTACATGTTGACAGTGTGAAGAACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAT 1200  
DB |||||

DB 1338 GAGTACATGTTGACAGTGTGAAGAACATTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAC 1397  
DB |||||

QY 1201 AAAAAAGAGTATATCTGGCTACTGATGATCTTCTTGTGTTAAAGAGGAGCAAGCAAG 1260  
DB |||||

DB 1398 AAAAAAGAGTATATTTGGCCACAGATGACCCCTTCTTTTAAAGAGGAGCAAAACAAAG 1457  
DB |||||

QY 1261 TACTCCAAATTAAGATTTATAGTGATACTCTATTTCTTGGTTCAGCTGGACTACACAT 1320  
DB |||||

DB 1458 TACCCCAATTAAGATTTATAGTGATACTCTATTTCTTGGTTCAGCTGGACTGCACAT 1517  
DB |||||

QY 1321 CGGTACACAGAAATTCACATTTGGGGTGTGATCTGATATACATTTCTCTCACAGCT 1380  
DB |||||

DB 1518 CGATACACAGAAATTCACATTTGGGGTGTGATCTGATATACATTTCTCTCACAGCA 1577  
DB |||||

QY 1381 GACTTCTAGTGTGTACTTTTTCATCCAGCTGTCTCGGGTGTCTTATGAATCATGCAA 1440  
DB |||||

DB 1578 GACTTCTAGTGTGTACTTTTTCATCCAGCTGTCTCGAGTGTCTTATGAATTTATGCAA 1637  
DB |||||

QY 1441 ACCCTGATCTGATGCTCTGGAACTTCCATCTTTTGGATGACATCTACTATTTTGA 1500  
DB |||||

DB 1638 ACATCATCTGATGCTCTGCAAACTTCCATTTCTTTTAGATGACATCTACTATTTTGGG 1697  
DB |||||

QY 1501 GGCCAAATCCCAACATCAGATTGCTGTTTATCTCCACAAAACCTCCAACTGGAAGAGAA 1560  
DB |||||

DB 1698 GGCCAAATCCCAACATCAGATTGCTGTTTATGCTCACCACCCCAACCTGAGATGAA 1757  
DB |||||

QY 1561 ATTCCAAATGGAACCTGAGATATCATTTGGTGTGGCTGGAAACCAATGGGATGTTTCT 1620  
DB |||||

DB 1758 ATTCCCAATGGAACCTGAGATATCATTTGGTGTGGCTGGAAATCATTTGGATGCTATTCT 1817  
DB |||||

1621	Qy	AAAGGTATCAACAGAGAAACTTGGAAAAACAGGCTTATATCCCTCCTCAAGATCCGAGAG	1680
1818	Db	AAAGGTGTCAACAGAGAAATTGGGAAGACGGGGCTATATCCCTCCTCAAGATTCGAGAG	1877
1681	Qy	AAGATAGAAAACAGTCAAGTATCCCATATCTCCTGAAGCTGAAAAATA	1727
1878	Db	AAGATAGAAAACGGTCAAGTACCCACATATCTCCTGAGCTGAGAAAAATA	1924

### RESULT 3

```

US-08-913-805A-1
; Sequence 1, Application US/08913805A
; Patent No. 6054304
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036

```

```

/ 2000
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3+ Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
/ SOFTWARE: WordPerfect 6.1 Windows
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/913.805A
/ FILING DATE: 7 JAN 1998
/

```

CLASSIFICATION: 435  
PRIOR APPLICATION DATE: PCT/JP97/00171  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 193260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796

```

/ DEBRAX; 202-423-0790 1:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1728 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA

```

Query Match 85.0%; Score 1469.4; DB 3; Length 1728;  
Best Local Similarity 90.7%; Pred. No. 0;  
Matches 1566; Conservative 0; Mismatches 161; Indels 0;

[illegible]

Db 1141 GAATACACGGTGCACGTTGAAGAAGACTTTCAGCTCTTGTCTGCAGAAATCAAGTGGAT 1200  
Qy 1201 AAAAAAGAGTATATCTGCTACTCATGATCTACTCTTTGTTAAAGGAGCGCAAGCAAG 1260  
Db 1201 AAAAAAGGGTATTTGGCCACAGATGACCTGCTTTGTTAAAGAGCGCAAAACAAG 1260  
Qy 1261 TACTCCAAATTAATGAATTTATTAGTGATAACTCTATTCTTGGTCTGAGTGGACTACACAAT 1320  
Db 1261 TACCCCAAGTATGAATTTATTAGTGATAACTCTATTCTTGGTCTGAGTGGACTACATAAT 1320  
Qy 1321 CGGTACACAGAAATTCATCTCGGGGTGTGATCTGGGATATACATTTCTCTCAGGCT 1380  
Db 1321 CGATATACAGAAATTCATCTCGGGGTGTGATCTGGGATATACATTTCTCTCCAGGCA 1380  
Qy 1381 GACTTCTAGTGTGTAATTTTTCATCCAGGCTGTCTCGGGTGTGATTAATGAATCATGCAA 1440  
Db 1381 GACTTCTAGTGTGTAATTTTTCATCCAGGCTGTCTCGGGTGTGATTAATGAATCATGCAA 1440  
Qy 1441 ACCCTGCAATCTGATCTCTGCGAACTTCCATCTTTTGGATGACATCTACTATTTTGA 1500  
Db 1441 GCGCTGCATCTGATCGCTCTGCGAACTTCCATCTTTTGGATGACATCTACTATTTTGA 1500  
Qy 1501 GGCACAAATGCCCAATCAGATTCCTGTTTATCTCTCAAACTCGAACTGAAGAGAA 1560  
Db 1501 GGCACAAATGCCCAATCAGATTCCTGTTTATCTCTCAAACTCGAACTGAAGAGAA 1560  
Qy 1561 ATTCCAATGGAACTGGAGATATCATCTGTGTGGCTGGAACCATTTGGGATGGTTATTCT 1620  
Db 1561 ATCCCAATGGAACTGGAGATATTAATTTGGTGTGGCTGGAATCATCTGGATGGCTATCT 1620  
Qy 1621 AAAGGTATCAACAGAAAACTTTGAAAAACAGGCTTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 1621 AAAGGTATCAACAGAAAACTGGAGAGGAGGCGCTATATCTCTCAAAAGTCCGAGAG 1680  
Qy 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAACTGAAATA 1727  
Db 1681 AAGATAGAAACAGTCAAGTATCCCAATATCTCTGAACTGAAATA 1727

## RESULT 4

US-09-442-629-1

; Sequence 1, Application US/09442629

; Patent No. 6291219

; GENERAL INFORMATION:

; APPLICANT: TANIGUCHI, Naoyuki

; UOZUMI, Naofumi

; SHIBA, Tetsuo

; YANAGIDANI, Shusaku

; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kenyon &amp; Kenyon

; STREET: 1025 Connecticut Avenue, N.W., Suite 600

; CITY: Washington

; STATE: DC

; COUNTRY: US

; ZIP: 20036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3+ Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2

; SOFTWARE: WordPerfect 6.1 Windows

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/442,629

; FILING DATE: 18-NO. 6291219-1999

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/913,805A

; FILING DATE: 7 JAN 1998

; APPLICATION NUMBER: PCT/JP97/00171

; FILING DATE: 23 JAN 1997

; APPLICATION NUMBER: JP 192260

; FILING DATE: 22 JUL 1996

; APPLICATION NUMBER: JP 162813

; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-439-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1728 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-442-629-1

Query Match 85.0%; Score 1469.4; DB 3; Length 1728;

Best Local Similarity 90.7%; Pred. No. 0;

Matches 1566; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 1 ATCGGGCGATGAGCTGTTCTCTGCGGTTGGATTATGCTCATTTCTTTTGGCGGGGACC 60

Db 1 ATCGGGCGATGAGCTGTTCTGCGGTTGGATTATGCTCATTTCTTTTGGCGGGGACC 60

Qy 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCCCTGATCATCTCCAGC 120

Db 61 TTGCTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCCCTGATCATCTCCAGC 120

Qy 121 AGAGAACTCTCCAAAGATCTTGGAAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180

Db 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGGAACGCTTAAACAGCAAAATGAAGACTTG 180

Qy 181 AGCGGAATGGCTGAGTCTCTCCGAATACCAAGGCGCCATTGACAGGGGACAGCTACA 240

Db 181 AGGAGATGGCTGAGTCTCTCCGAATACCAAGGCGCCATTGATCAGGGGCGAGCTTCA 240

Qy 241 GGAAGAGTCCGTGTTTATGAAGAAACAGCTTGTATGAAGCCAAAGAACAGATTTGAAAATTTAC 300

Db 241 GGAAGAGTCCGTGTTTATGAAGAAACAGCTTGTATGAAGCCAAAGAACAGATTTGAAAATTTAT 300

Qy 301 AAGAAACAACTAGAAATGGTCTCGGGAAGATCATGAATCTTAAGAGAGAGATTGAA 360

Db 301 AAGAAACAACTAGAAATGGTCTCGGGAAGATCATGAATCTTAAGAGAGAGATTGAA 360

Qy 361 AATGAGCTAAAGAGCTCTGTTTCTCAAAAGCGAATCTGAAGAAATTAAGCAATTTA 420

Db 361 AATGAGCTAAAGAGCTCTGTTTCTCAAAAGCGAATCTGAAGAAATTAAGCAATTTA 420

Qy 421 GAAGGAAATGAATCTCCAAAGACATGCAATTTCTTGGATTTTAGGACACCATGAA 480

Db 421 GAAGGAAATGAATCTCCAAAGACATGCAATTTCTTGGATTTTAGGACACCATGAA 480

Qy 481 AGGTCTATCATGACAGATCTTATCTACCTAGTCAAAAGAGATGGAGCGGGATTTGGCGT 540

Db 481 AGGTCTATCATGACAGATCTTATCTACCTAGTCAAAAGAGATGGAGCGGGATTTGGCGT 540

Qy 541 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATTAACATCTCCAGAT 600

Db 541 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATTAACATCTCCAGAT 600

Qy 601 CCTAAGGACTGCAGCAAAAGCCAGGAAGCTGGTGTGAACATCAATTAAGGCTGTGGCTAT 660

Db 601 CCTAAGGACTGCAGCAAAAGCCAGGAAGCTGGTGTGAACATCAATTAAGGCTGTGGCTAT 660

Qy 661 GGTTGTCAACTCCATGAGCTGGTGTCTACTGTTTCTATGATTTGTCACCCAGCAACA 720

Db 661 GGTTGTCAACTCCATGAGCTGGTGTCTACTGTTTCTATGATTTGTCACCCAGCAACA 720

QY	721	CTCATCTTGGAAATCTCAGAAATTGGCGCTATGCTACTGCTGGATGGGAGACTGTTT	780
DB	721	CTCGCTTGGAACTTCACAATTGGCGCTACGCTACTGGGGATGGGAAACTGTGTTT	780
QY	781	CCTGTAAGTGAGACATGTAACAGACAGATCTGGGCTCTCCACTGGACACTGGT	840
DB	781	CCTGTAAGTGAGACGTCACACAGACAGATCTGGCAGCTCCACTGGACATTTGGT	840
QY	841	GTAATGACAAAACATTTCAAGTGGTCGAGCTCCCAATCTGACAGACCTCCATCT	900
DB	841	GTAAGACAAAATGTTTCAAGTGGTTGAGCTCCCAATCTGACAGAGTTCATCTCT	900
QY	901	CCTCTTACTTACCACTGGCTGTTCCAGAAGACCTTGCAGACCGACTCTTAAGAGT	960
DB	901	CCTCCATATTACCCCTGGCTGTCCAGAGACCTTGCAGATCGACTTGTACAGTCC	960
QY	961	GGTGACCTCGAGTGGTGGGTGCCAGTTGTCAAATCTTTGATTCCTCACAACT	1020
DB	961	GGTGATCTCGAGTGGTGGGTATCCAGTTGTCAAAGTACTTGTTCGCCACAACCC	1020
QY	1021	TGCTGGAAAGGAATAGAGAGGACCAAGAGCTTGGCTTCAAACATCCAGTTATT	1080
DB	1021	TGCTGGAAAGGAATAGAGAGGACCAAGAGCTAGGCTTCAAACATCCAGTTATT	1080
QY	1081	GGAGTCCATCTCAGACGACACAGAAAGTGGGAACAGAAGCAGCTTCCACCCATCG	1140
DB	1081	GGAGTCCATGTTAGACGACAGCAAGTGGGAGCGGAGCAGCTTCCATCCCATGAG	1140
QY	1141	GAGTACATGTAACAGTTGAAGAACATTTTCAGCTTCTCGACGCAAGATGCAAGTGG	1200
DB	1141	GAATACACGGTGACAGTTGAAGAAGACTTTTCAGCTTCTCGCAGATGCAAGTGG	1200
QY	1201	AAAAAAGAGTATCTGGCTACTGATGATCTTACTTTGTTAAAGGAGGCAAGACAA	1260
DB	1201	AAAAAAGGGGTGATTTGGCCACAGATGACCTTGTGTTAAAGGAGGCAAAACAA	1260
QY	1261	TACTCCAAATTTAGTTAGTGAATCTTATTTCTTGGTCAGCTGACATACACAAT	1320
DB	1261	TACCCAGTTAGATTTATTTAGTGAATCTTATCTTGGTCAGCTGACATACATA	1320
QY	1321	CGGTACACAGAAATTCACCTCGGGGTGATCTCTGGATATACATTTCTCTCAGGCT	1380
DB	1321	CGATATACAGAAATTCACCTCGGGGTGATCTCTGGATATACACTTCTCTCCAGCA	1380
QY	1381	GACTTCTAGTGTACTTTTTCATCCAGGTCTGTCGGGTGCTTATGAATCATGCA	1440
DB	1381	GACTTCTAGTGTACTTTTTCATCCAGGTCTGTTAGAGTGTGCTATGAATCATGCA	1440
QY	1441	ACCTGTCATCTGATGCTCTGCGAACTTCCATTTTGGATGACATCTACTATTTTGG	1500
DB	1441	CGCTGTCATCTGATGCTCTGCGAACTTCCGTTCTTGGATGACATCTACTATTTG	1500
QY	1501	GGCCAAATGCCCAATCAGATGTGTTTATCTCTCACAAACCTCGAATGAAGAGAA	1560
DB	1501	GGCCAAATGCCCAACCAATTTGCCATTTATCTCTCACAACTCGAATGAAGAGAA	1560
QY	1561	ATTCCANTGAACTTGAGATATCATTTGGTGTGGCTGGAACCATTTGGGATGGTTCT	1620
DB	1561	ATCCCATGAACTTGAGATATTTATTTGGTGTGGCTGGAATCATCTGGGATGGCTCT	1620
QY	1621	AAAGGTATCAACAGAAAATTTGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAG	1680
DB	1621	AAAGGTATTAACAGAAAATTTGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAG	1680
QY	1681	AAGATAGAAACAGTCAAGTATCCCATATCTCTGAAGCTGAAATA	1727
DB	1681	AAGATAGAAACAGTCAAGTATCCCATATCTCTGAAGCTGCAAGTA	1727

Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-3300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9pt-F18  
PS-08-232-463-14







## TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:

## INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

LENGTH: 2116

TYPE: Nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Human

US-09-000-041A-1

Query Match 2.2%; Score 37.6; DB 3; Length 2116;

Best Local Similarity 49.6%; Pred. No. 0.18;

Matches 126; Conservative 0; Mismatches 124; Indels 4; Gaps 1;

Qy 280 AAAGAACAGATTGAAATTTACAAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAA 339

Db 1507 AATGAACAATGGACCTTGGCTATCCCAAGCCATCAGCTCTGGAAGGGATCCCTGA 1566

Qy 340 ATCTTAAGAGGAGGATTAAGAAATGGAGCTAAAGAGCTCTGTTTTTTTCTACAAAGCGAA 399

Db 1567 ATCTCTCAGGAGCATTTGTACACAAAGAAATGGCTTTAGCTATTCTTCTACAAAGGAAA 1626

Qy 400 CT----GAAGAAATTAAGCATTTAGAGGAAATGAATCCCAAGACATGAGATGAAT 455

Db 1627 GGAGTATTGAAATTTCAACACACCATGATCTCAAGGTAGAACCTGGACATCCAAGATCCAT 1686

Qy 456 TCTTTTGATTTAGGACACCATGAAGGTCTATCATGACAGATCATACTACCTCAGTCA 515

Db 1687 CCTAAGATTTTATGGCTGTGATGGACCAACAGACAGATTAAGAGGACACAGCCC 1746

Qy 516 AACGATGGAGCAG 529

Db 1747 ACCAGATGATGAG 1760

## RESULT 9

US-09-134-001C-2083

Sequence 2083, Application US/09134001C

Patent No. 6380370

## GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 2083

LENGTH: 1596

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-2083

Query Match

Best Local Similarity 63.3%; Pred. No. 0.2;

Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 237 TACGAGAGAGTCCTGTTTGAAGAAACAGCTTCTTAAGGCCAAGAACAGATTGAAAA 296

Db 177 TAAGAGACAGACAATATATAAAGAAAAAATTTACTTGAAGGCAAGAAAGAGAACCAAT 236

Qy 297 TTACAGAAACAAAGCTAGAAATGGTCTGGG 326

Db 237 TTTAAAGACACAGCTGAGATGACTGCG 266

## RESULT 10

US-08-933-750C-54/c

Sequence 54, Application US/089333750C

Patent No. 5932442

## GENERAL INFORMATION:

APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer L.

APPLICANT: Bandman, Olga

APPLICANT: Shah, Purvi

APPLICANT: Au-Young, Janice

APPLICANT: Yue, Henry

APPLICANT: Guegler, Karl J.

TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

NUMBER OF SEQUENCES: 98

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/933,750C

FILING DATE: September 23, 1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0356 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 1750 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: HMCINOT01

CLONE: 9476

US-08-933-750C-54

Query Match 2.1%; Score 36.8; DB 2; Length 1750;

Best Local Similarity 54.4%; Pred. No. 0.29;

Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 304 AACCAAGCTAGAAATGTCTCGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAAAAT 363

Db 1718 AAAAACTTAGAGAGAGAGAGAGAGAGATCAAAATTAACCTATGAAGATTTATAGAAAAC 1659

Qy 364 GGAGCTAAGAGCTCTGGTTTTTCTCAAAAGCGAAGCTGAAGAAATTAAGCATTTAGAA 423

Db 1658 ACACGAGTAGGAGGTTCTTTTCAAGCAACCAACAGACAGTTACGTAACCTAAAA 1599

Qy 424 GGAATGAACTCCAAA 439

Db 1598 AGAAATGCTTTCCACA 1583

## RESULT 11



Query Match	2.1%;	Score 36;	DB 4;	Length 4517;
Best Local Similarity	45.4%;	Pred. No. 1;		
Matches 167;	Conservative 0;	Mismatches 200;	Indels 1;	Gaps 1;
Qy	183	GCGAATGGCTGAGTCTCTCCGAATACAGAAAGGCCCATTCACACAGGGGACAGCTACAGG	242	
Db	4127	GCTCATGGTAGAGGTAGCAGTGCACCAAGATCACACCACTGAACTCCAGCCAGGGCCACA	4068	
Qy	243	AAGAGTCCGTGTTTTAGAGAACAGCTTTGTTAAGCCAAAGAACAGATTGAAAAATTACAA	302	
Db	4067	GAACGAGAGTCTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA	4008	
Qy	303	GAACAACAGCTAGAAATGCTCTGGGGAAAGGATCATGAAATCTTTAAGAAGGAGGATTGAAAA	362	
Db	4007	GGAAGGAAGAAA	3948	
Qy	363	TGGAGCTAAAGAGCTCTGGT- TTTTTCTACAAAGCGAACTGAAGAAATTTAAAGCATTTAG	421	
Db	3947	GGAAGTTGAGTGCCTATGTTATTTTATTTAGACAAAGTAGACTTCAGAACACACAGAAAAATTATC	3888	
Qy	422	AAGGAATGAACTCCAAGACATGCAGATGAAATTTCTTTTGGATTTAGGACACCATGAAA	481	
Db	3887	ACAGATAAGAAGGACATTACATAATGATAGAGAGTAAATGCAACCAATTAAGACATAATA	3828	
Qy	482	GGTCTATCATGCACAGATCTACTACCTCAGTCAACAGATGGAGCAGGGGATTGGCGTG	541	
Db	3827	CTAAACATGTATATATCTAATAATAGAGCGCTAAATAATATATGAAGCAAAAGCTGACAGAA	3768	
Qy	542	AAAAAGAG	549	
Db	3767	CTAAAGAG	3760	

```

RESULT 14
US-09-776-976-5/c
; Sequence 5, Application US/09776976
; Patent No. 6566332
; GENERAL INFORMATION:
; APPLICANT: Fruebis, Joachim
; APPLICANT: Erickson, Mary Ruth
; APPLICANT: Yen, Frances
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
; FILE REFERENCE: 76.US4.REG
; CURRENT APPLICATION NUMBER: US/09/776,976
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/758,055
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: US 60/176,228
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/198,087
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: US 60/299,881
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 5
; LENGTH: 4517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-776-976-5

```

Qy	303	GAAACAAGCTAGAAAATGCTCGGGGAAGGATCATGAAATCTTTAAGAAGGAGATTGAAAA	362
Db	4007	GGAAGGAAACAGAGACAGAGAGAAAAGGAAGGAAGAAAAGGAAGGAAGGAAGAAA	3948
Qy	363	TGGAGCTAAAGAGCTCTGGT-TTTTTCTACAAAGCGAACCTGAAGAAATTTAAAGCATTTAG	421
Db	3947	GGAAGTTGAGTCGGTATGTATTTTTTAGACAAAGTAGACTTTCAACACACAGAAAATTATC	3888
Qy	422	AAGGAAATGAACGCCAAGAAGCATGCAGATGAAAATCTTTTGAGTTTAGGACACATGAAA	481
Db	3887	ACAGATAAAGAGAGCACATTACATAATGATAAGAGAGTAATATGCACCATAAAGACATAATA	3828
Qy	482	GGTCTATCATGACAGATCTATACCTACCTCAACAGATGAGCAGCGGGATTGGCGTG	541
Db	3827	CTAAACATGTATATATCTAATAATAGAGCGCTAAAAATATATGAAGCAAAGCTGCACAGAA	3768
Qy	542	AAAAAGAG	549
Db	3767	CTAAGAG	3760

RESULT 15  
US-09-909-547-5/C  
; Sequence 5, Application US/0909547  
; Patent No. 6579852  
; GENERAL INFORMATION:  
; APPLICANT: Fruebis, Joachim  
; APPLICANT: Erickson, Mary Ruth  
; APPLICANT: Yen, Frances  
; APPLICANT: Bhaini, Bernard  
; TITLE OF INVENTION: OBG Globular Head and Uses Thereof for Decreasing Body  
; FILE REFERENCE: 76.US6.CIP  
; CURRENT APPLICATION NUMBER: US/09/909,547  
; CURRENT FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: US 09/776,976  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 09/758,055  
; PRIOR FILING DATE: 2001-01-10  
; PRIOR APPLICATION NUMBER: US 60/299,881  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: US 60/198,087  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: US 60/176,228  
; PRIOR FILING DATE: 2000-01-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5  
; LENGTH: 4517  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-909-547-5

	Query Match	2.1%;	Score 36;	DB 4;	Length 4517;
	Best Local Similarity	45.4%;	Pred. No. 1;		
	Matches 167;	Conservative 0;	Mismatches 200;	Indels 1;	Gaps 1;
Qy	183	CGGNATGGCTGAGCTCTCTCCGAATACAGAGAAGGCCCATTTACCCAGGGGACAGCTACAGG	242		
Db	4127	GCTCATGGTAGAGGTAGCTGAGCGCCAAAGATCACACCCTGAACTCCAGCCAGCGGCCACCA	4068		
Qy	243	AAGAGTCCGTGTTTTAGAAGAACAGCTTGTTTAAGGCCAAAGAACAGATTGAAAATTACAA	302		
Db	4067	GAACGAGAGTCTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA	4008		
Qy	303	GAACAAGCTAGAAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTTGAAA	362		
Db	4007	GGAAGGAAAGAGAGAGAGAGAGAGAAAGAAAGGGAAGGAAGAAAGAGGAGGAGAAAGAAA	3948		
Qy	363	TGGAGCTAAAGAGCTCTGGT-TTTTCTCAAGACCGAACTGAGAAATTTAAGCATTTTAG	421		
Db	3947	GGAAGTTGAGTGGCGTATGTTATTTTATAGACAAAGTAGACTTTCAGACACAGAAAATTTATC	3888		





Db 61 TTGTTATTTATATAGTGGTCAATTTGGTTCGAGATAATGACCAACCTGATCACTCCAGC 120  
Qy 121 AGAGAACTCTCCAGATTTCTTCAAGCTTTGAACCTTAAACACACAAAATGAAGACTTG 180  
Db 121 AGAGAACTCTCCAGATTTCTTCAAGCTTTGAACCTTAAACACACAAAATGAAGACTTG 180  
Qy 181 AGCGAAATGGCTGAGTCTCTCCGAATACACAGAAAGCCCACTTTGACAGGGACAGCTACA 240  
Db 181 AGCGAAATGGCTGAGTCTCTCCGAATACACAGAAAGCCCACTTTGACAGGGACAGCTACA 240  
Qy 241 GGAAGAGTCCGTTGTTTGAAGAACAGACTTTGTTAAGGCGCAAGAACAGATGAAAATTAC 300  
Db 241 GGAAGAGTCCGTTGTTTGAAGAACAGACTTTGTTAAGGCGCAAGAACAGATGAAAATTAC 300  
Qy 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360  
Db 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTAAAGAGGAGGATTGAA 360  
Qy 361 AATGAGAGCTAAGAGCTCTGTTGTTTCTTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
Db 361 AATGAGAGCTAAGAGCTCTGTTGTTTCTTCAAAAGCGAACTGAAGAAATTTAAAGCATTTA 420  
Qy 421 GAAGAAATGAATCTCAAGACATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Db 421 GAAGAAATGAATCTCAAGACATGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Qy 481 AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGGAGAGGGGATGGCGT 540  
Db 481 AGGTCTATCATGACAGATCTATCTACCTCAGTCAACAGATGGAGAGGGGATGGCGT 540  
Qy 541 GAAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAACATCTCCAGAAAT 600  
Db 541 GAAAAGAGGCGCAAGATCTGACAGAGTGGTCCAGCGGAGAAATTAACATCTCCAGAAAT 600  
Qy 601 CTAAGAGACTGAGCAAGACCGAGAGTGGTGTGTAAATCAATCAATCAATCAATCAATCAAT 660  
Db 601 CTAAGAGACTGAGCAAGACCGAGAGTGGTGTGTAAATCAATCAATCAATCAATCAATCAAT 660  
Qy 661 GGTCTCACTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Db 661 GGTCTCACTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
Qy 721 CTCATCTTGGAACTCTCAGAAATGGCGCTATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Db 721 CTCATCTTGGAACTCTCAGAAATGGCGCTATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCCAGGTGAA 840  
Db 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTCCAGGTGAA 840  
Qy 841 GTAAATGACAAAACATTTCAAGTGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Db 841 GTAAATGACAAAACATTTCAAGTGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Qy 901 CCTCTCTTACTTACCACTGGCTGCTCCAGAAAGCTTTGACAGACCGACTCTTAAGAGTCCAT 960  
Db 901 CCTCTCTTACTTACCACTGGCTGCTCCAGAAAGCTTTGACAGACCGACTCTTAAGAGTCCAT 960  
Qy 961 GGTGACCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
Db 961 GGTGACCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020  
Qy 1021 TGGCTGGAAGGAATAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Db 1021 TGGCTGGAAGGAATAGAAAGCCACCAAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Qy 1081 GGAGTCCATGTGACAGCGACAGACAAAGTGGGAAACAGAGCGCTTCCACCCCATCGAG 1140  
Db 1081 GGAGTCCATGTGACAGCGACAGACAAAGTGGGAAACAGAGCGCTTCCACCCCATCGAG 1140  
Qy 1141 GAGTACATGGTACAGTGTGAAGAAATTTTCAAGTCTTCTGACAGCAGAAATGCAAGTGGAT 1200

Db 1141 GAGTACATGGTACACAGTTTGAAGAACATTTTTCAGCTTCTCGCACGCAAGATGCAAGTGGAT 1200  
Qy 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTCTACTTTGTTTAAAGGAGGCAAAAGCAAAG 1260  
Db 1201 AAAAAAGAGATATATCTGGCTACTGATGATCTCTACTTTGTTTAAAGGAGGCAAAAGCAAAG 1260  
Qy 1261 TACTCCAAATTAATGAATTTATAGTGAATACTCTATTTCTTGGTTCAGTGGGACTACAAAT 1320  
Db 1261 TACTCCAAATTAATGAATTTATAGTGAATACTCTATTTCTTGGTTCAGTGGGACTACAAAT 1320  
Qy 1321 CGGTACACAGAAAATTCACCTTCGGGGTGTGATCTGGATATACACTTTCTCTCACAGGCT 1380  
Db 1321 CGGTACACAGAAAATTCACCTTCGGGGTGTGATCTGGATATACACTTTCTCTCACAGGCT 1380  
Qy 1381 GACTTTCTAGTGTACTTTTTTCAATCCAGGTCTGTTCGGGTTGCTTATGAAATCATGCAA 1440  
Db 1381 GACTTTCTAGTGTACTTTTTTCAATCCAGGTCTGTTCGGGTTGCTTATGAAATCATGCAA 1440  
Qy 1441 ACCCTGATCTGATGCTCTGCAACTTCCTGATCTCTTGGATGACATCTACTATTTTGA 1500  
Db 1441 ACCCTGATCTGATGCTCTGCAACTTCCTGATCTCTTGGATGACATCTACTATTTTGA 1500  
Qy 1501 GGCCAAAATGCCCAATCAGATTTGCTTTATCTCACAACCTCGAACTGAAAGAGAA 1560  
Db 1501 GGCCAAAATGCCCAATCAGATTTGCTTTATCTCACAACCTCGAACTGAAAGAGAA 1560  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1620  
Db 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1620  
Qy 1621 AAAGGTATCAACAGAAAACCTTGGAAAACAGGGCTTATATCCCTCTACAAAGTCCGAGAG 1680  
Db 1621 AAAGGTATCAACAGAAAACCTTGGAAAACAGGGCTTATATCCCTCTACAAAGTCCGAGAG 1680  
Qy 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAATAG 1728  
Db 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCGAAGCTGAAAATAG 1728

## RESULT 2

US-09-971-773-1  
; Sequence 1, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANA1  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; PRIOR FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2008  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-1

Query Match: 91.7%; Score 1584; DB 11; Length 2008;  
Best Local Similarity 94.8%; Pred. No. 0;  
Matches 1638; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 1 ATCGGGGATGGACTGGTTCCTGCGTGGATTATGCTCATCTTTTTCCTGGGGGACC 60



Db 100 ATCGGGGATGACTGGTTCCTGGCGTGGATATGCTCATTTCTTTTGGCTCGGGGACC 159  
Qy 61 TTGTTATTTTATATAGGTGGTTCATTTGGTTCGAGATATGACCCCTCGATCACTCCAGC 120  
Db 160 TTATTTGTTTATATAGGTGGTTCATTTGGTTCGAGATATGACCCCTCGATCACTTCAGC 219  
Qy 121 AGAGAACTCTCCAGATTTCTTGCAAGCTTGAAACGCTTTAAACAGCAAAATGAAGACTTG 180  
Db 220 AGAGAACTCTCCAGATTTCTTGCAAGCTTGAAACGCTTTAAACAGCAAAATGAAGACTTG 279  
Qy 181 AGGGAATGGCTGAGTCTCTCGAATACAGAGGCCCTTATGACCGGGGACAGCTTACA 240  
Db 280 AGGGAATGGCTGAGTCTCTCGAATACAGAGGCCCTTATGATCAGGGGACAGCTACA 339  
Qy 241 GGAAGAGTCCGTGTTTGAAGAGACAGCTTGTGAAGGCCAAAGAACAGATTTGAAAATTTAC 300  
Db 340 GGAAGAGTCCGTGTTTGAAGAGACAGCTTGTGAAGGCCAAAGAACAGATTTGAAAATTTAC 399  
Qy 301 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTTAAGGAAGGAGATTGAA 360  
Db 400 AAGAAACAAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTTAAGGAAGGAGATTGAA 459  
Qy 361 AATGAGCTAAAGAGCTCTGTTTCTTCAAAAGCGAACTGAAGAAATTAAGCATTTA 420  
Db 460 AATGAGCTAAAGAGCTCTGTTTCTTCAAAAGCGAACTGAAGAAATTAAGCATTTA 519  
Qy 421 GAAGGAATGACTCCAAAGCATGACAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Db 520 GAAGGAATGACTCCAAAGCATGACAGATGAAATCTTTTGGATTTAGGACACCATGAA 579  
Qy 481 AGGTCTATCATGACAGATCTATCTACTCTCAAGACAGATGAGAGCGGATTTGGCGT 540  
Db 580 AGGTCTATCATGACAGATCTATCTACTCTCAAGACAGATGAGAGCGGATTTGGCGG 639  
Qy 541 GAAAGAGGCCAAAGATCTGACAGAGCTGCTCAGCGGAGAAATTAACATATCTCCAGAT 600  
Db 640 GAAAGAGGCCAAAGATCTGACAGAGCTGCTCAGCGGAGAAATTAACATATCTCCAGAT 699  
Qy 601 CCTAAGGACTGACAGAAAGCCAGGAGCTGCTGATCAATCAATTAAGGCTTGGCTAT 660  
Db 700 CCTAAGGACTGACAGAAAGCCAGGAGCTGCTGATCAATCAATTAAGGCTTGGCTAT 759  
Qy 661 GTTGTCAACTCCATCAGCTGCTTACTGTTTCTATGATTTGCTTATGGCAACCCAGCAACA 720  
Db 760 GGATGTCACCTCCATCATGCTGTTTACTGCTTCATGATTTGTTATGGCAACCCAGCAACA 819  
Qy 721 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTGA 780  
Db 820 CTCATCTTTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGAGACTGTGTTTGA 879  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCAGTGGACACTGTGTCAGGTGAA 840  
Db 880 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCAGTGGACACTGTGTCAGGTGAA 939  
Qy 841 GTAAATGACAAAACATTCAGTGTGCTGAGCTGCCCTTTAGAGAGCTCCATCTCCG 900  
Db 940 GTGAAGHACAAAATGTTCAAGTGTGCTGAGCTGCCCTTTAGAGAGCTCCATCTCCG 999  
Qy 901 CCTCTTACTTACCCTGCTGTTCCAGAGACCTTTGACAGAGCTCCATCTTAAAGAGTCCAT 960  
Db 1000 CCTCTTACTTACCCTGCTGTTCCAGAGACCTTTGACAGAGCTCCATCTTCCAGAGTCCAT 1059  
Qy 961 GGTGACCTCGAGTGTGGTGTGCTCCAGATTTGTTCAAAATCTGATTTGCTCCCAACCT 1020  
Db 1060 GGTGATCTCGAGTGTGGTGTGCTCCAGATTTGTTCAAAATCTGATTTGCTCCCAACCT 1119  
Qy 1021 TGGCTGGAAGGAAATGAGACAGCTCCAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Db 1120 TGGCTGGAAGGAAATGAGACAGCTCCAGAGCTTGGCTTCAACATCCAGTTATT 1179  
Qy 1081 GGAGTCCATGTACAGCGCACAGACAAAAGTGGGAACAGAGCGCTTCCACCCCATCGAG 1140

Db 1180 GGATCCCATGTACAGACGACTGACAAAAGTGGGAAACAGAGCAGCGCTTCCATCCCATGAG 1239  
Qy 1141 GAGTACATGGTACACGCTTGAAGAAACATTTTTCAGCTTCTCGCACGACAGATCAAGTGGAT 1200  
Db 1240 GAATACATGGTACACGCTTGAAGAAACATTTTTCAGCTTCTCGAACGAGATCAAGTGGAT 1299  
Qy 1201 AAAAAAGAGTATATCTGGCTACATGATGATCTTCTTGTGTTTAAAGGAGGCAAGCAAAAG 1260  
Db 1300 AAAAAAGAGTATCTGGCCACTGATGACCCCTTCTTGTGTTTAAAGGAGGCAAGCAAAAG 1359  
Qy 1261 TACTCCAAATATGAATTTATAGTATGAATCTTATTTCTTGGTCACTGAGTACACAAAT 1320  
Db 1360 TACTCCAAATATGAATTTATAGTATGAATCTTATTTCTTGGTCACTGAGTACACAAAT 1419  
Qy 1321 CGGTACACAGAAAATTCATCTTGGGGTGTGATCTGGATATACATTTTCTCTCACAGGCT 1380  
Db 1420 CGATACACAGAAAATTCATCTTGGGGTGTGATCTGGATATACATTTTCTCTCCAGGCT 1479  
Qy 1381 GACTTTCTAGTGTGTAATTTTTCATCCAGGCTCTGTGGGTTGCTTATGAAATCATGCAA 1440  
Db 1480 GACTTCTGTGTGTAATTTTTCATCCAGGCTCTGTAGGTTGCTTATGAAATCATGCAA 1539  
Qy 1441 ACCCTGATCTGATGCTCTCGAACTTCCATTTCTTTGATGATCACTACTATTTTGA 1500  
Db 1540 AACTGATCTGATGCTCTCGAACTTCCATTTCTTTAGATGATCTACTATTTTGA 1599  
Qy 1501 GGCCAAAATGCCAATCAGATTTGCTTTATCTCACAACCTCGAACTGAAGAGAA 1560  
Db 1600 GGCCAAAATGCCAATCAGATTTGCTTTATCTCACAACCTCGAACTGAAGAGAA 1659  
Qy 1561 ATTCCAATGGAACCTGGAGATATCATTTGGTGTGCTGGAAACCATTTGGATGTTTATCT 1620  
Db 1660 ATCCCATGGAACCTGGAGATATCATTTGGTGTGCTGGAAACCATTTGATGTTTACTCT 1719  
Qy 1621 AAAGGTATCAACAGAAACCTTGGAAACAGGCTTATATCCCTCTTACAAAGTCCGAGAG 1680  
Db 1720 AAAGGTATCAACAGAAACCTTGGAAACAGGCTTATATCCCTCTTACAAAGTCCGAGAG 1779  
Qy 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCAGCTGAAATAG 1728  
Db 1780 AAGATAGAAACAGTCAAGTATCCCTACATATCTCAGCTGAAATAG 1827

RESULT 3  
US-09-839-136-9  
; Sequence 9, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: FCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 2100  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (198)....(1925)  
US-09-839-136-9

Query Match 86.7%; Score 1498.2; DB 9; Length 2100;  
Best Local Similarity 91.7%; Pred.No. 0;  
Matches 1584; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 1 ATGCGGCGATGAGTGGTTCCTGGCGTTGGATTAATGCTCATCTCTTTTTCCTGGGGACC 60  
Db 198 ATGCGGCGATGAGTGGTTCCTGGCGTTGGATTAATGCTCATCTCTTTTTCCTGGGGACC 257  
QY 61 TTGTTATTTTATATAGTGGTCATTTGGTTGGATTAATGACCCCTGATCAGTCCAGC 120  
Db 258 TTGCTGTTTATATAGTGGTCATTTGGTTGGATTAATGACCCCTGATCAGTCCAGC 317  
QY 121 AGAGAACTCTCCAGATTTCTGCAAGCTTGAACGTTTAAACACGCAAAATGAAGACTTG 180  
Db 318 CGAGAACTGTCCAGATTTCTGCAAGCTTGAACGTTTAAACACGCAAAATGAAGACTTG 377  
QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCTCATTTGACAGGGGACAGCTACA 240  
Db 378 AGGCGAATGGCGCAATCTCTCCGATACCAAGAGCCCTCATTTGATCAGGGGCCAGCTATA 437  
QY 241 GGAAGAGTCGTTGTTTGAAGAACAGCTTTGTTTGAAGGCCAAAGAACAGATTTGAATAATAC 300  
Db 438 GGAAGAGTACGCGTTTGAAGAGAGAGCTTTGTTTGAAGGCCAAAGAACAGATTTGAATAATAC 497  
QY 301 AAGAAACAGCTAGAAATGGTCTGGGGAAGGATCATGAATCTTAAGAGAGGAGGATTCGAA 360  
Db 498 AAGAAACAGACCAAGAAATGGTCTGGGGAAGGATCATGAATCTTAAGAGAGGAGGATTCGAA 557  
QY 361 AATGAGAGTAAAGAGCTCTGGTTTTTTCTACAAAGCGAACTGAAGAAATTAAGAGCATTTA 420  
Db 558 AATGAGAGTAAAGAGCTCTGGTTTTTTCTACAGAGTGAATTAAGAGAACTTA 617  
QY 421 GAAGGAATGAATCTCAAGACATCGAGATGAATTTCTTTGGATTTAGGACACCATGAA 480  
Db 618 GAAGGAATGAATCTCAAGACATCGAGATGAATTTCTTTGGATTTAGGACATCATGAA 677  
QY 481 AGGCTATCATGACAGATCTACTACTCTCAGTCAACAGATGAGAGAGGAGTTGGCGT 540  
Db 678 AGGCTATTAATACCGGATCTATACTACTCTCAGTCAACAGATGAGAGAGGAGTTGGCGG 737  
QY 541 GAAAGAGGCGCAAGATCTGACAGAGTGTGTCAGCGGAGAAATTAACATATCTCCAGAA 600  
Db 738 GAAAGAGGCGCAAGATCTGACAGAACTGGTTTCAAGCGGAGAAATTAACATATCTCCAGAA 797  
QY 601 CTTAAGGAGTGAAGAACAGGAGAGTGGTGTGTATCATCAATTAAGAGGCTGGCTAT 660  
Db 798 CTTAAGGAGTGAAGAACAGGAGAGTGGTGTGTATCATCAATTAAGAGGCTGGCTAT 857  
QY 661 GGTGTGCACTCACTGCTGTCTACTGTTTCTATGTTTATGTCACCCAGCGAACA 720  
Db 858 GGTGTGCACTCACTGCTGTCTACTGTTTCTATGTTTATGTCACCCAGCGAACA 917  
QY 721 CTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGAGAGCTGTGTTAGA 780  
Db 918 CTCATCTTGGAAATCTCAGAAATGGCGCTATGCTACTGTTGGATGGAGAGCTGTGTTAGA 977  
QY 781 CTTGTAAGTGAACATGTACAGACAGATCTGGCGCTCTCCACTGGAACCTGGTCAAGTGAA 840  
Db 978 CTTGTAAGTGAACATGTACAGACAGATCTGGCGCTCTCCACTGGAACCTGGTCAAGTGAA 1037  
QY 841 GTAAATGCAAAACATTAAGTGTGCTGAGCTCCCATTTGATGAGAGCTCCATCTCGG 900  
Db 1038 GTAAAGGCAAAATTTCAAGTGTGCTGAGCTCCCATTTGATGAGAGCTCCATCTCGG 1097  
QY 901 CCTCTCTACTTACCACTGGCTCTTCCAGAAAGCTTTGACAGCGGAGCTCTTAAGAGTCCAT 960  
Db 1098 CCTCTCTACTTACCACTGGCTCTTCCAGAAAGCTTTGACAGCGGAGCTCTTAAGAGTCCAT 1157  
QY 961 GGTGACCTGCAAGTGTGGTGTGTCAGGATTTGTCATTAATCTGATTCGTCACAACT 1020  
Db 1158 GGTGACCTGCAAGTGTGGTGTGTCAGGATTTGTCATTAATCTGATTCGTCACAACT 1217  
QY 1021 TGGCTGGAAGGAATGAAGAGCCCAAGAGAGCTTGGCTTCAACATCCAGTTAT 1080  
Db 1218 TGGCTGGAAGGAATGAAGAGCCCAAGAGAGCTTGGCTTCAACATCCAGTTAT 1277

QY 1081 GGAGTCCATGTCAGACGACAGACAAAGTGGACAGAGAGCTTCCACCCCATCGAG 1140  
Db 1278 GGAGTCCATGTCAGACGACAGACAAAGTGGACAGAGAGTCCATCCCATCGAG 1337  
QY 1141 GAGTACATGGTACACGCTTGAAGAACATTTTTCAGCTTCTCGCACGAGAAATGCAAGTGGAT 1200  
Db 1338 GAGTACATGGTACACGCTTGAAGAACATTTTTCAGCTTCTCGCACGAGAAATGCAAGTGGAT 1397  
QY 1201 AAAAAAGAGTATATCGGCTACATGATGATCTCTACTTTGTTTAAAGAGGAGGCAAGCAAG 1260  
Db 1398 AAAAAAGAGTATATCGGCTACATGATGATCTCTACTTTTAAAGAGGAGGCAAGCAAG 1457  
QY 1261 TACTCCAAATATGAATTTATAGTGAATCTCTATTTTCTTGGTCTGAGTGGAGTACACAA 1320  
Db 1458 TACCCCAATATGAATTTATAGTGAATCTCTATTTTCTTGGTCTGAGTGGAGTACACAA 1517  
QY 1321 CGGTACACAGAAATTCACCTTCGGGGTGTATCTCGATATACATCTTCTCTCAGAGCT 1380  
Db 1518 CGATACACAGAAATTCACCTTCGGGGTGTATCTCGATATACATCTTCTCTCAGAGCT 1577  
QY 1381 GACTTCTAGTGTACTTTTTTCAATCCAGGTCTGTCTGGTGTGTTATGAAATCATGCA 1440  
Db 1578 GACTTCTAGTGTACTTTTTTCAATCCAGGTCTGTCTGGTGTGTTATGAAATCATGCA 1637  
QY 1441 ACCGTGATCTGATGCTCTCGAACTTCCATTTCTTGGTGTGTTATGAAATCATGCA 1500  
Db 1638 ACATACATCTGATGCTCTGCAAACTTCCATTTCTTGGTGTGTTATGAAATCATGCA 1697  
QY 1501 GGCCAAATGCCCAATTCAGATTTGTTTATCTCACAACCTCGAACTGAAAGAGGAA 1560  
Db 1698 GGCCAAATGCCCAATTCAGATTTGTTTATCTCACAACCTCGAACTGAAAGAGGAA 1757  
QY 1561 ATTCCAATGGAACCTGGAGATATCAITTTGGTGTGCTGAAACCTGAAATGTTTATCT 1620  
Db 1758 ATTCCAATGGAACCTGGAGATATCAITTTGGTGTGCTGAAACCTGAAATGTTTATCT 1817  
QY 1621 AAAGGTATCAACAGAAACTTGGAAACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1680  
Db 1818 AAAGGTATCAACAGAAACTTGGAAACAGGCTTATATCTCTCTCAAAAGTCCGAGAG 1877  
QY 1681 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAATA 1727  
Db 1878 AAGATAGAAACAGTCAAGTATCCACATATCTCTGAAGCTGAAATA 1924

## RESULT 4

US-10-106-698-381  
; Sequence 381, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: P8005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: Patent In Ver. 3.0  
; SEQ ID NO 381  
; LENGTH: 3007  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2984)..(2984)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (2988)..(2988)

OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (3002)...(3002)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-10-106-698-381

Query Match 86.2%; Score 1490; DB 15; Length 3007;

Best Local Similarity 91.7%; Pred. No. 0;

Matches 1585; Conservative 1; Mismatches 141; Indels 1; Gaps 1;

```
QY 1 ATGGGGGATGGAGCTGGTCTCTGGGTTGGATTATGCTCATCTTTTGGCTCGGGGACC 60
DB 420 ATGGGGGATGGAGCTGGTCTCTGGGTTGGATTATGCTCATCTTTTGGCTCGGGGACC 479
QY 61 TTGTTATTTATATAGGTGGTCTTGGTTCGAGATATGACCCCTGATCACTCCAGC 120
DB 480 TTGCTGTTTTATATAGGTGGTCTTGGTTCGAGATATGACCCCTGATCACTTAGC 539
QY 121 AGAGAACTCTCAAGATTCTTGAAGCTTGAACGCTTTAAACAGCAAAATGAAGACTTG 180
DB 540 CGAGAACTGTCCAAGATTCTGCAAGCTTGAACGCTTTAAACAGCAAAATGAAGACTTG 599
QY 181 AGGGAATGGGTG-AGTCTCTCGAATACAGAGAGGCCCAATGACAGGGGACAGCTAC 239
DB 600 AGGGAATGGGCAAAATCTCTCGGATACAGAGAGGCCCTATGATCAGGGGGCCAGCTAT 659
QY 240 AGGAAGCTCGGTGTTTGAAGAACAGCTTGTAAAGCCCAAGAACAGATTGAAATTA 299
DB 660 AGGAAGCTCGGTGTTTGAAGAACAGCTTGTAAAGCCCAAGAACAGATTGAAATTA 719
QY 300 CAAGAAA CAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTTAAAGAGGAGATTGA 359
DB 720 CAAGAAA CAAGCTAGAAATGGTCTGGGGAAGGATCATGAAATCTTTAAAGAGGAGATTGA 779
QY 360 AAATGGAGCTAAAGAGCTCTGGTCTTCTCAAGAGCAAGCTTGAAGAAATTAAGCACTT 419
DB 780 AAATGGAGCTAAAGAGCTCTGGTCTTCTCAAGAGCAAGCTTGAAGAAATTAAGCACTT 839
QY 420 AGAAGAAATGAACCTCAAGACATGACAGATGAATCTTTTGGATTAGGACACCATGA 479
DB 840 AGAAGAAATGAACCTCAAGACATGACAGATGAATCTTTTGGATTAGGACACCATGA 899
QY 480 AAGTCTATCATGACAGATCTATACCTCAGTCAAAACAGATGAGAGGGGATTTGGCG 539
DB 900 AAGTCTATATGACGGATCTATACCTCAGTCAAGACAGATGAGAGGGGATTTGGCG 959
QY 540 TGAAGAGAGGCAAGATCTGACAGAGCTGTCCAGCGGAGATTAATATCTCCAGAA 599
DB 960 TGAAGAGAGGCAAGATCTGACAGAGCTGTCCAGCGGAGATTAATATCTCCAGAA 1019
QY 600 TCCTAAGAGCTGACAGCAAGGAGAGCTGGTGTAACTCAATCAATCAAGGCTGTGGCTA 659
DB 1020 TCCTAAGAGCTGACAGCAAGGAGAGCTGGTGTAACTCAATCAATCAAGGCTGTGGCTA 1079
QY 660 TGGTGTCAACTCCATCACTGCTGTCTACTGTTTCAATGATTGCTTATGCAACCCAGCGAAC 719
DB 1080 TGGTGTCACTGCTCACTGCTGTCTACTGTTTCAATGATTGCTTATGCAACCCAGCGAAC 1139
QY 720 ACTCATCTTGAATCTCAAGATTGGCGCTATGCTACTGTTGATGAGAGAGCTGTGTTTGA 779
DB 1140 ACTCATCTTGAATCTCAAGATTGGCGCTATGCTACTGTTGATGAGAGAGCTGTGTTTGA 1199
QY 780 ACCTGTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
DB 1200 GCCTGTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1259
QY 840 AGTAATGACAAAAATCTCAAGTGGTCCAGCTCCCAATTTGACAGAGCTCCATCTCCG 899
DB 1260 AGTGAAGACAAAAATCTCAAGTGGTCCAGCTCCCAATTTGACAGAGCTCCATCTCCG 1319
QY 900 GCCTCTTACTTACCACTGGCTGTTTCCAGAGAGACCTTGCAGACCGAATCTTAAAGAGTCCA 959
DB 1320 TCCTCCATATTTACCTTGGCTGTACCAAGAGACCTTGCAGATCGACTTGTACGAGTGCA 1379
```

```
QY 960 TGTGACCCCTGCAGTGTGGTGGGTGTCCAGTTTGTCAAATACTTGTATCGTCCACAACC 1019
DB 1380 TGTGACCCCTGCAGTGTGGTGGGTGTCTCAGTTTGTCAAATACTTGTATCGTCCACAACC 1439
QY 1020 TTGGCTGAAAAAGGAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1079
DB 1440 TTGGCTGAAAAAGGAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATCCAGTTAT 1499
QY 1080 TGGAGTCCATGTCAAGCGCAGCAAAAGTGGGAACAGAGCTGCTTCCATCCCATGGA 1139
DB 1500 TGGAGTCCATGTCAAGCGCAGCAAAAGTGGGAACAGAGCTGCTTCCATCCCATGGA 1559
QY 1140 GGAGTACATGATCAACGTTTGAAGAACATTTTTCAGCTTCTCGCACGCAAGATGCAAGTGA 1199
DB 1560 AGAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1619
QY 1200 TAAAAAAGAGTATATCTGGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1259
DB 1620 CAAAAAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679
QY 1260 GTACTCCAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1319
DB 1680 GTACTCCAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1739
QY 1320 TCGGTACACAGAAATTCACCTTCGGGGTGTGATCCTCGATATACACTTCTCTCACAGGC 1379
DB 1740 TCGGTACACAGAAATTCACCTTCGGGGTGTGATCCTCGATATACACTTCTCTCACAGGC 1799
QY 1380 TGACTTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1439
DB 1800 AGACTTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1859
QY 1440 AACCTCCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1499
DB 1860 AACCTCCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1919
QY 1500 AGGCAAAATGCCCAATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1559
DB 1920 AGGCAAAATGCCCAATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1979
QY 1560 AATTCGAATGAACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1619
DB 1980 AATTCGAATGAACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2039
QY 1620 TAAAGGTATCAACAGAAATCTTGAAGAACAGGCTTATATCCCTCCTCAAGTCCGAGA 1679
DB 2040 TAAAGGTATCAACAGAAATCTTGAAGAACAGGCTTATATCCCTCCTCAAGTCCGAGA 2099
QY 1680 GAAGATAGAAACAGTCAAGTATCCCATATCTTGAAGCTGAAAAATA 1727
DB 2100 GAAGATAGAAACAGTCAAGTATCCCATATCTTGAAGCTGAAAAATA 2147
```

## RESULT 5

US-09-839-136-1  
; Sequence 1, Application US/09839136  
; Patent No. US200200816941  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1

```
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Pig
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1728)
US-09-839-136-1

Query Match      84.9%; Score 1467.8; DB 9; Length 1728;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1565; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1 ATGCGGCGATGACGTGGTCTCGGCTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60
DB 1 ATGCGGCGATGACGTGGTCTCGGCTGGATTATGCTCATTTCTTTTGGCTGGGGGACC 60

QY 61 TTGTTATTTTATATAGTGGTCAATTTGGTTCGAGATATGACCACTGATCACTCCAGC 120
DB 61 TTGCTATTTTATATAGTGGTCAATTTGGTTCGAGATATGACCACTGATCACTCCAGC 120

QY 121 AGAGAACTCTCCAAAGTCTTGCAAGCTTGAAAGCTTTAAACAGCAAAATGAAGACTTG 180
DB 121 CGAGAACTCTCCAAAGTCTTGCAAGCTTGAAAGCTTTAAACAGCAAAATGAAGACTTG 180

QY 181 AGGCGAATGGCTGAGTCTCTCCGAATACAGAAAGCCCATTTGACCGGGGACAGCTACA 240
DB 181 AGGCGAATGGCTGAGTCTCTCCGAATACAGAAAGCCCATTTGATCAGGGGCGAGCTTCA 240

QY 241 GGAAGAGTCTGTTTGAAGAAACAGCTTTGTTAAGGCGCAAGACAGATTTGAAATTTAC 300
DB 241 GGAAGAGTCTGTTTGAAGAAACAGCTTTGTTAAGGCGCAAGACAGATTTGAAATTTAT 300

QY 301 AAGAACAAGCTAGAAATGGTCTGGGAGAGATCATGAATCTTAAAGAGAGAGATTGAA 360
DB 301 AAGAACAACAATTAAGATGGTCCAGGAGAGATCATGAATCTTAAAGAGAGAGATTGAA 360

QY 361 AATGAGCTTAAAGAGCTCTGTTTCTTCAAGCGAACTGAAGAAATTAAGCATTTA 420
DB 361 AATGAGCTTAAAGAGCTCTGTTTCTTCAAGCGAACTGAAGAAATTAAGCATTTA 420

QY 421 GAAGGAATGAATCTCAAGACATCGAGATGAATTTCTTGGATTTAGGACACCATGAA 480
DB 421 GAAGGAATGAATCTCAAGACATCGAGATGAATTTCTTATCAGATTTGGGACATCATGAA 480

QY 481 AGGTCTATCATCAGATCTATCTACTACTCTCAGTCAAGAGAGAGAGAGAGAGAGAG 540
DB 481 AGGTCTATTAAGCGATCTATCTACTACTCTCAGTCAAGAGAGAGAGAGAGAGAGAG 540

QY 541 GAAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 600
DB 541 GAAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 600

QY 601 CTAAGGAGTGCAGAAAGCCAGGAGAGCTGGTGTGTAATCAATTAAGGCTGGGCTAT 660
DB 601 CCAAGGAGTGCAGAAAGCCAGGAGAGCTGGTGTGTAATCAATTAAGGCTGGGCTAT 660

QY 661 GTTGTCTCACTCATCAGCTGCTTACTGTTTCTGATTTGCTTATGGCAACCCAGCGAACA 720
DB 661 GGTCTGTAGCTTCATCATGTAGTGTACTGCTTTTATGATTTGATGTCACCCAGCGAACA 720

QY 721 CTCATCTTTGGAATCTCAGAAATTTGGCGTATGCTACTGGTGGATGGGAGACTGTGTTAGA 780
DB 721 CTCGCTTTGGAATCTCAGAAATTTGGCGTACGCTACTGGGGAATGGGAACTGTGTTAGA 780

QY 781 CTTGTAAGTGCAGATATGACAGAGATCTGCGCTCTCCACTGGGACACTGTGTCAGTAAA 840
DB 781 CTTGTAAGTGCAGATATGACAGAGATCTGCGCAGCTCTCCACTGGGACACTGTGTCAGTAAA 840

QY 841 GTAAATGACAAAAACATTCAGTGTGCGAGCTCCCATTTGTAGACAGCTCTCATCTCGG 900
DB 841 GTAAAGGACAAAATGTTTCAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCATCTCGT 900

QY 901 CCTCTCTTACTTACCAGTGGCTGTTCCAGAAAGACCTTGCAGACCGGACTCCTAAGAGTCCAT 960
```

```
DB 901 CCTCATATTACCCCTGGCTGTCCAGAAAGACCTTGAGATCGACTTGTACGAGTCCAT 960
QY 961 GGTGACCTGTCAGTGTGGTGTCCAGTGTTCCTCAAAATCTTGTTCGATTCGTCACAACT 1020
DB 961 GGTGATCTGTCAGTGTGGTGTTCCTCAAAATCTTGTTCGATTCGTCACAACT 1020
QY 1021 TGGCTGAAAAAGGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATCTTGTTCAT 1080
DB 1021 TGGCTGAAAAAGGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAAATCTTGTTCAT 1080
QY 1081 GGAGTTCATGTCACAGCAGACAGAAAGTGGAAAGAGAGCTTCCACCCATCCAG 1140
DB 1081 GGAGTTCATGTCACAGCAGACAGAAAGTGGAAAGAGAGCTTCCATCCCATTCAG 1140
QY 1141 GAGTACATGGTACACAGCTTGAAGAACTTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAT 1200
DB 1141 GAATACAGCTGTCAGCTTGAAGAACTTTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAT 1200
QY 1201 AAAAAAGAGATATCTGGCTACTGATGATCTTCTTGTGTTTAAAGAGGCAAAAGCAAAAG 1260
DB 1201 AAAAAAGAGGATATTTGGCCACAGATGACCTCTGTTTAAAGAGGCAAAAGCAAAAG 1260
QY 1261 TACTCCAAATTAAGATTTTATAGTAACTCTATTTCTTGGTTCAGCTGGGACTACAAAT 1320
DB 1261 TACCCAGTATGATGATTTTATAGTAACTCTATTTCTTGGTTCAGCTGGGACTACAAAT 1320
QY 1321 CGGTACACAGAAATTCACCTTGGGGTGTGATCTGATATACACTTTCTCTCACAGGCT 1380
DB 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCTGATATACACTTTCTCTCCAGGCA 1380
QY 1381 GACTTCTAGTGTACTTTTTCATCCAGCTGTGTCGGGTGCTTATGAAATCATGCAA 1440
DB 1381 GACTTCTAGTGTACTTTTTCATCCAGCTGTGTCGGGTGCTTATGAAATCATGCAA 1440
QY 1441 ACCCTGATCTGATGCTCTGCGAACTTCCTTCTTGGTTCAGCTACATCTACTATTTTGA 1500
DB 1441 GCGTGTGATCTGATGCTCTGCGAACTTCCTTCTTGGTTCAGCTACATCTACTATTTTGA 1500
QY 1501 GGCCAAATGCCCAATTCAGATTTGCTTATCTCACAACCTCGAACTGAAAGAGAA 1560
DB 1501 GGCCAAATGCCCAATTCAGATTTGCTTATCTCACAACCTCGAACTGAAAGAGAA 1560
QY 1561 ATTCCAATGGAACTCGAGATATCATTTGCTGTGCTGGAACCAATGGGATGGTATTTCT 1620
DB 1561 ATCCCATGGAACTCGAGATATTTATTTGGTGTGCTGGAATCACTGGGATGGCTATCT 1620
QY 1621 AAAGGTATCAACAGAAACTTGGAAACAGAGCTTATCTCTCTACAAAGTCCGAGAG 1680
DB 1621 AAAGGTGTAAACAGAAACTTGGAAAGAGCGGGCTTATATCTCTCTACAAAGTCCGAGAG 1680
QY 1681 AAGATAGAAACAGTCAAGATATCCACATATCTCTGAACTGAAAAATA 1727
DB 1681 AAGATAGAAACAGTCAAGATATCCACATATCTCTGAACTGAAAAATA 1727
```

## RESULT 6

```
US-09-971-773-7
; Sequence 7, Application US/09971773
; Publication No. US20030115614A1
; GENERAL INFORMATION:
; APPLICANT: Yutaka KANDA
; APPLICANT: Mitsuo SATOH
; APPLICANT: Kazuyasu NAKAMURA
; APPLICANT: Kazuhisa UCHIDA
; APPLICANT: Toyohide SHINKAWA
; APPLICANT: Naoko YAMANE
; APPLICANT: Motoo YAMASAKI
; APPLICANT: No. US20030115614A1uo HANAI
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971,773
; CURRENT FILING DATE: 2002-08-30
```

; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 979  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-09-971-773-7

Query Match 53.9%; Score 931; DB 11; Length 979;

Best Local Similarity 96.9%; Pred. No. 1.9e-262;

Matches 949; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy 720 ACTCATCTTGAATCTCAGAATTCGAGATTCGCTATGCTACTGGTGGATGGGAGACTGTTTAG 779
Db 1 ACTCATCTTGAATCTCAGAATTCGAGATTCGCTATGCTACTGGTGGATGGGAGACTGTTTAG 60

Qy 780 ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA 839
Db 61 ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA 120

Qy 840 AGTAAATGACAAAAATTCAGAGTTCGAGTCCCAATTTGTAGACAGCTCCATCCTCG 899
Db 121 AGTAAATGACAAAAATTCAGAGTTCGAGTCCCAATTTGTAGACAGCTCCATCCTCG 180

Qy 900 GCCTCTTACTTACACTGGCTGTTCCAGAGACCTTTCAGACAGCTCCCTAAGATCCA 959
Db 181 GCCTCTTACTTACCACCTGGCTGTTCCAGAGACCTTTCAGACAGCTCCCTAAGATCCA 240

Qy 960 TGGTGACCTCGAGTGTGGTGTCCAGTTCCTCAATTTGATTCGTCACAAACC 1019
Db 241 TGGTGACCTCGAGTGTGGTGTCCAGTTCCTCAATTTGATTCGTCACAAACC 300

Qy 1020 TTGGCTGAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTAT 1079
Db 301 TTGGCTGAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTAT 360

Qy 1080 TGGAGTCCATGTACAGCAGCAGACAAAGTGGGAACAGAGAGCTTCCACCCATCGA 1139
Db 361 TGGAGTCCATGTACAGCAGCAGACAAAGTGGGAACAGAGAGCTTCCACCCATCGA 420

Qy 1140 GGAGTACATGTACAGTGTGAAGACATTTTCAGTCTTCGACGACGAGTGAAGTGA 1199
Db 421 GGAGTACATGTACAGTGTGAAGACATTTTCAGTCTTCGACGACGAGTGAAGTGA 480

Qy 1200 TAAAAAAGAGTATATCTGGCTACTGTATGATCTCTTGTAAAGGAGGCAAGACAAA 1259
Db 481 TAAAAAAGAGTATATCTGGCTACTGTATGATCTCTTGTAAAGGAGGCAAGACAAA 540

Qy 1260 GTACTCCAAATATGAATTTATAGTGAATCTCTATTTCTTGGTCAGTGAACACAA 1319
Db 541 GTACTCCAAATATGAATTTATAGTGAATCTCTATTTCTTGGTCAGTGAACACAA 600

Qy 1320 TCGGTACACAGAAATTCACCTTCGGGTGTATCTGTGATATACACTTCTCTCAGGC 1379
Db 601 TCGGTACACAGAAATTCACCTTCGGGTGTATCTGTGATATACACTTCTCTCAGGC 660

Qy 1380 TGACTTTCTAGTGTACTTTTTCATCCAGGTCTGTGGGTGCTGTATGAAATCATGCA 1439
Db 661 TGACTTTCTAGTGTACTTTTTCATCCAGGTCTGTGGGTGCTGTATGAAATCATGCA 720

Qy 1440 AACCTGTGATCTGTGCTCTCGGAATCTCCATTTCTTGGATGACATCTACTATTTGG 1499
Db 721 AACCTGTGATCTGTGCTCTCGGAATCTCCATTTCTTGGATGACATCTACTATTTGG 780

Qy 1500 AGGCCAAATGCCCAATCAGATGCTGTTTATTCCTCAAACTCGAATCGAAGGAGGA 1559
Db 781 AGGCCAAATGCCCAATCAGATGCTGTTTATTCCTCAAACTCGAATCGAAGGAGGA 840

Qy 1560 AATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGCGAAACCATTTGGGATGTTTTC 1619
```

```
Db 841 AATTCCAATGGAACCTGGAGATATCATTTGGTGTGGCTGGAACCATTTGGGATGTTTTC 900
Qy 1620 TAAAGGTATCAACAGAAACTTTGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGA 1679
Db 901 TAAAGGTGTCAACAGAAACTTTGAAAAACAGGCTTATATCCCTCTACAAAGTCCGAGA 960
Qy 1680 GAAGATAGAAACAGTCAAG 1698
Db 961 GAAGATAGAAACGCTCAAG 979
```

# RESULT 7

US-09-971-773-6  
; Sequence 6, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTI-BODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 979  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-09-971-773-6

Query Match 51.7%; Score 894.2; DB 11; Length 979;

Best Local Similarity 94.6%; Pred. No. 1.2e-251;

Matches 926; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy 720 ACTCATCTTGAATCTCAGAATTCGCGCTATGCTACTGGTGGATGGGAGACTGTTTAG 779
Db 1 ACTCATCTTGAATCTCAGAATTCGCGCTATGCTACTGGAGGATGGGAGACTGTTTAG 60

Qy 780 ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA 839
Db 61 ACCTGTAAGTCAGACATGTACAGACAGATCTGGCTCTCCACTGCACACTGGTCAGGTGA 120

Qy 840 AGTAAATGACAAAAATTCAGAGTTCGAGTCCCAATTTGTAGACAGCTCCATCCTCG 899
Db 121 AGTAAATGACAAAAATTCAGAGTTCGAGTCCCAATTTGTAGACAGCTCCATCCTCG 180

Qy 900 GCCTCTTACTTACCATCTGGCTGTTCCAGAGACCTTTCAGACAGCTCCCTAAGATCCA 959
Db 181 GCCTCTTACTTACCATCTGGCTGTTCCAGAGACCTTTCAGAGACCTCCCTGAGAGTCCA 240

Qy 960 TGGTGACCTCGAGTGTGGTGTCCAGTTCCTCAATTTGATTCGTCACAAACC 1019
Db 241 TGGTGACCTCGAGTGTGGTGTCCAGTTCCTCAATTTGATTCGTCACAAACC 300

Qy 1020 TTGGCTGAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTAT 1079
Db 301 TTGGCTGAAAGGAAATAGAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTAT 360

Qy 1080 TGGAGTCCATGTACAGCAGCAGACAAAGTGGGAACAGAGAGCTTCCACCCATCGA 1139
Db 361 TGGAGTCCATGTACAGCAGCAGACAAAGTGGGAACAGAGAGCTTCCACCCATCGA 420
```



```
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13359
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL09847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
US-09-864-761-13359

Query Match      15.6%; Score 269.2; DB 9; Length 503;
Best Local Similarity 89.8%; Pred. No. 3.4e-68;
Matches 289; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1406 CCCAGTCTGTGGGTTGTTATGAAATCATGCAAAACCTGCTGATGCTGCGA 1465
Db      |||||||
6 CCAAGTCTGTGAGTTGCTTATGAAATATGCAAACTACATCTGATGCTCTGCAA 65
QY 1466 ACTTCATCTTTGGATGACATCTACTATTTGGAGGCCAAATGCCACAAATCAGATTG 1525
Db      |||||||
66 ACTTCATCTTTAGATGACATCTACTATTTGGGGGCCAGAAATGCCCAAAATCAAAATTG 125
QY 1526 CTGTTATCTCTCACAACTGCAACTGAAGAGGAATTCCAATGGAACCTGGAGATCA 1585
Db      |||||||
126 CCAATTATGCTCACCACCCGCACTGAGATGAAATTTCCCATGGAACCTGGAGATCA 185
QY 1586 TTGGTGTGGTGAACCACTGGGATGTTTCTAAAGGTATCAACAGAAAACCTTGAA 1645
Db      |||||||
186 TTGGTGTGGTGAATCAATCTGGATGGCTATCTTAAGGTCTCAACAGGAATTTGGAA 245
QY 1646 AAACAGGCTTATATCCCTCTACAAAGTCCGAGAGAAGATGAAACAGTCAAGTATCCCA 1705
Db      |||||||
246 GGAAGGCTATATCCCTCTACAAAGTCCGAGAGAAGATGAAACAGTCAAGTATCCCA 305
QY 1706 CATATCTGAGCTGAAAATA 1727
Db      |||||||
306 CATATCTGAGCTGAGAAATA 327

RESULT 10
US-09-864-761-13292
; Sequence 13292, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13292
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ALL09847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
US-09-864-761-13292

Query Match      12.2%; Score 210.8; DB 9; Length 551;
Best Local Similarity 80.0%; Pred. No. 5.2e-51;
Matches 248; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 773 TGTTTAGACCTGTAAAGTGAGACATGTACAGACAGATCTGGGCTCTCCACTGGACACTGCTGT 832
Db      |||||||
116 TGTGTAATGGTGATATATAGGAATACCATGTGTATGATATATGTTCAATATGTT 175
QY 833 CAGGTGAAGTAAATGACAAAAAATTCAGTGGTGAGCTCCCATTTGTAGACAGCTCC 892
Db      |||||||
176 CAGGTGAAGTGAAGGACAAAAATGTTCAAGTGGTGAGCTTCCCATTTGTAGACAGCTCTTC 235
QY 893 ATCTCGGCTCTCTTACTTACACTGCTGTTCCAGAGAGCTTCCAGACGACTCCCTAA 952
Db      |||||||
236 ATCCCGGCTCTCCATATTTTACCTTGGCTGTACCAAGAGACTCGACATCGACTGTAC 295
QY 953 GAGTCCATGTTGACCTCGAGTGTGGTGGTGTCCAGTTTGTCAAATACTTGTATCGTC 1012
Db      |||||||
296 GAGTGCATGTTGACCTCGAGTGTGGTGGTGTCTCAGTTTGTCAAATACTTGTATCGTC 355
QY 1013 CACAACCTTGGCTGGAAAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATC 1072
Db      |||||||
356 CACAGCTTGGCTAGAAAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAAACATC 415
QY 1073 CAGTTATTGG 1082
Db      |||||||
416 CAGTTATTGG 425

RESULT 11
US-09-864-761-18181
; Sequence 18181, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
```



FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 18181  
LENGTH: 248  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL109847.1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9  
OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.50e+00  
OTHER INFORMATION: NT HIT: AF038280.1, EVALUE 0.00e+00  
OTHER INFORMATION: EST\_HUMAN HIT: AW387766.1, EVALUE 0.00e+00  
US-09-864-761-18181

Query Match 11.9%; Score 205.4; DB 9; Length 248;  
Best Local Similarity 89.5%; Pred. No. 1.2e-49;  
Matches 221; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 836 GTGAAGTAAATGACAAAACATTCAAGTGGTCGAGCTCCCATTTGTAGACAGCTTCATC 895  
DB 1 GTGAAGTGAAGGACAAAATGTTCAAGTGGTCGAGCTCCCATTTGTAGACAGCTTCATC 60  
QY 896 CTCGCGCTCCTTACTTACCACTGGCTGTCAGAGACCTTCACAGCCGACTCCTTAAG 955  
DB 61 CCCGCTCCTCATTTTACCCTTGGCTGTACCAAGAGACCTTCGACAGTGTGTACGAG 120

QY 956 TCCATGCTGACCTCCAGTGTGGTGGTCCCGAGTTGTCAAATACTTGTATTCGTCCAC 1015  
DB 121 TGCAATGCTGACCTCCAGTGTGGTGGTGTCTCAGTTGTCAAATACTTGTATTCGCCCCAC 180  
QY 1016 AACCTTGGCTGAAAAGAAATAGAAGAACCCACCAAGAGCTTGGCTTCAAACATCCAG 1075  
DB 181 AGCTTGGCTAGAAAAGAAATAGAAGAACCCACCAAGAGCTTGGCTTCAAACATCCAG 240  
QY 1076 TTATTGG 1082  
DB 241 TTATTGG 247

## RESULT 12

US-09-864-761-30284  
Sequence 30284, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 30284

LENGTH: 248

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AL109847.3

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4



```

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.50e+00
; OTHER INFORMATION: EST HUMAN HIT: AW387766.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: AF038280.1, EVALUE 0.00e+00
US-09-864-761-30284

```

Query Match	11.98;	Score	205.4;	DB	9;	Length	248;
Best Local Similarity	89.5%;	Pred. No.	1.2e-49;				
Matches	221;	Conservative	0;	Mismatches	26;	Indels	0;
Gaps	0;						
Qy	836	GTCAAGTAAATGACAAAACATTCAAGTGTGCGAGCTCCCATTTGTAGACAGAGCCTCCATC	895				
Db	1	GTCAAGTGAAGGACAAAATGTTCAAGTGTGCGAGCTTCCCATTTGTAGACAGCTTTCATC	60				
Qy	896	CTCGGCTCTTACTTACCACCTGGCTGTTCCAGAAGACCTTGCAGACCGACTCTCTAAGAG	955				
Db	61	CCGCTCTCCATATTTACCCTTGGCTGTACAGAAGACCTCGCAGATCGAGTGTGTACGAG	120				
Qy	956	TCCATGGTGACCTCGCAGTGTGGTGGTGTCGCCAGTTTGTCAAATACTTGTATGTCCTCAC	1015				
Db	121	TGCATGGTGACCTGCGAGTGTGGTGGTGCTCAGTTTGTCAAATACTTGTATCGGCCAC	180				
Qy	1016	AACTTTGGCTGGAAAAGGAAATAGAAGAAGCCACCAAGAAGCTTGGCTTCAAAACATCCAG	1075				
Db	181	AGCCTTGGCTAGAAAAGAAATAGAAGAAGCCACCAAGAAGCTTGGCTTCAAAACATCCAG	240				
Qy	1076	TTATTGG	1082				
Db	241	TTATTGG	247				

```

RESULT 13
US-09-971-773-3
; Sequence 3, Application US/09971773
; Publication No. US20030115614A1
; GENERAL INFORMATION:
; APPLICANT: Yutaka KANDA
; APPLICANT: Mitsuo SATOH
; APPLICANT: Kazuyasu NAKAMURA
; APPLICANT: Kazuhisa UCHIDA
; APPLICANT: Toyohide SHINKAWA
; APPLICANT: Naoko YAMANE
; APPLICANT: Motoo YAMASAKI
; APPLICANT: No. US20030115614A1uo HANAI
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971,773
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: JP 2000-308526
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/268,926
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 9196
; TYPE: DNA
; ORGANISM: Cricetulus griseus
US-09-971-773-3

```

```

Db      2669 AGAGAACTCTCCAGAATTTTCGAAGCTGGACCGCTTAAACAACAATAAGACTTG 2728
Qy      181 AGCGAATGCTGAGTCTCTCCG 203
Db      2729 AGGAGAATGCTGAGTCTCTCCG 2751

RESULT 14
US-10-125-968-274
; Sequence 274, Application US/10125968
; Publication No. US20030215805A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Palermo, Adam
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; APPLICANT: Elias, Josh
; APPLICANT: Mertens, Maureen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-032
; CURRENT APPLICATION NUMBER: US/10/125,968
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,163
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 1417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 204, 266, 320, 328, 402, 424, 462, 470
; OTHER INFORMATION: n = A,T,C or G
US-10-125-968-274

```

RESULT 15  
US-09-864-761-1421  
; Sequence 1421, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; FILE REFERENCE: Aeonica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 1421  
;; LENGTH: 384  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL109847.1  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6  
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.6  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9  
US-09-864-761-1421

Query Match 7.3%; Score 126.8; DB 9; Length 384;  
Best Local Similarity 73.9%; Pred. No. 2e-26;  
Matches 161; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
Qy 773 TGTTTAGACCTTAAGTGAGACATCTGGCTCTCCACTGGACACTGGT 832  
Db |||||  
167 TCTTCAATGGTGGATGATAGGAATACCATGTGTAATGATTATATGTTCAATATTGT 226  
Qy 833 CAGGTGAAGTAATGACAAAACATTCAGTGGTCGAGTCCCATTTGTAGACAGCCTCC 892  
Db |||||  
227 CAGGTGAAGTAAGGACAAAATGTTCAAGTGGTCGAGCTTCCCATTTGTAGACAGTCTTC 286

Qy 893 ATCTCGCCTCCTTACTTACCACCTGGCTGTTCCAGAGACCTTGCAGACCGACTCTAA 952  
Db |||||  
287 ATCCCCGCTCTCCCATATTTACCTTGGCTGTACCAAGACCTCGCAGATCGACTGTGAC 346  
Qy 953 GAGTCCATGGTGACCTGCGAGTGTGGTGGGTGTCCAG 990  
Db |||||  
347 GAGTGCATGGTGACCTGCGAGTGTGGTGGGTGTCTCAG 384

Search completed: February 2, 2004, 19:03:05  
Job time : 616.522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 2, 2004, 11:54:20 ; Search time 3711.31 Seconds  
(without alignments)  
11316.242 Million cell updates/sec

Title: US-09-971-773-2  
Perfect score: 1728  
Sequence: 1 atggggcagtgactggttc.....atctgaagctgaaaaatag 1728

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0  
Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estmu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_htc.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_htc.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_ges\_hum.\*
- 18: em\_ges\_inv.\*
- 19: em\_ges\_pln.\*
- 20: em\_ges\_vrt.\*
- 21: em\_ges\_fun.\*
- 22: em\_ges\_mam.\*
- 23: em\_ges\_mus.\*
- 24: em\_ges\_pro.\*
- 25: em\_ges\_rod.\*
- 26: em\_ges\_phg.\*
- 27: em\_ges\_vrl.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	100.0	2766	11 AK051811	AK051811 Mus muscu
2	1728	100.0	3052	11 AK048520	AK048520 Mus muscu
3	913.2	52.8	937	13 BQ892101	BQ892101 AGENCOURT
4	801.6	46.4	898	14 CB194828	CB194828 AGENCOURT

5	787.6	45.6	1066	12	BM928521	BM928521 AGENCOURT
6	747.2	43.2	940	13	BQ653329	BQ653329 AGENCOURT
7	746.8	43.2	900	13	BX452321	BX452321
8	735.2	42.5	886	13	BQ423860	BQ423860 AGENCOURT
9	725	42.0	923	13	BX429921	BX429921
10	720.4	41.7	855	13	BUI84296	BUI84296 AGENCOURT
11	704.6	40.8	937	13	BUS13162	BUS13162 AGENCOURT
12	696.8	40.3	1027	13	BX450010	BX450010
13	687.4	39.8	832	9	AUI24128	AUI24128
14	660.2	38.2	702	12	BG973221	BG973221
15	656	38.0	656	13	BQ552954	BQ552954
16	627.2	36.3	752	13	BQ604588	BQ604588
17	627.2	36.3	752	13	BX414472	BX414472
18	620.2	35.9	965	12	B1546364	B1546364
19	615.2	35.6	790	14	CB989769	CB989769
20	613	35.5	613	14	CA539475	CA539475
21	608.8	35.2	692	14	CB105611	CB105611
22	608.8	35.2	692	14	CB105621	CB105621
23	593	34.3	805	14	CA489662	CA489662
24	592.8	34.3	717	14	CB322023	CB322023
25	576.4	33.4	769	10	BG501191	BG501191
26	568.2	32.9	573	9	AI122135	AI122135
27	567.8	32.9	936	10	BE893107	BE893107
28	562.8	32.6	646	14	CB105920	CB105920
29	559	32.3	715	12	BG771085	BG771085
30	557.2	32.2	900	13	BQ731540	BQ731540
31	554.2	32.1	938	13	BX433236	BX433236
32	551	31.9	970	13	BQ926824	BQ926824
33	550.6	31.9	637	14	CB105617	CB105617
34	548.2	31.7	731	13	BX456265	BX456265
35	542	31.4	622	14	CB105602	CB105602
36	541	31.3	621	14	CB105873	CB105873
37	538	31.1	618	14	CB105618	CB105618
38	535.2	31.0	653	10	BG619839	BG619839
39	531.8	30.8	607	14	CB105608	CB105608
40	531.8	30.8	607	14	CB105882	CB105882
41	531.8	30.8	608	14	CB105918	CB105918
42	530.2	30.7	607	14	CB105899	CB105899
43	529.8	30.7	534	10	BE690071	BE690071
44	524.2	30.3	610	14	CB105867	CB105867
45	512.4	29.7	586	14	CB105886	CB105886

ALIGNMENTS

RESULT 1

AK051811

LOCUS

DEFINITION

AK051811 2766 bp mRNA linear HTC 05-DEC-2002  
Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length  
enriched library, clone:D330003K17 product: fucosyltransferase 8,  
full insert sequence.

ACCESSION

AK051811.1 GI:26342229

VERSION

HTC; CAP trapper.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

PUBMED

11042159

REFERENCE  
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

REFERENCE  
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, K., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Mateu, Y., Nikaide, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzaletti, J., Mombaerts, P., Nordone, P.,  
King, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,  
Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

TITLE  
JOURNAL  
MEDLINE  
PUBMED

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2766)

REFERENCE  
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitho, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp.  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

TITLE  
JOURNAL

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

## FEATURES

## source

1. 2766  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:D230003K17"  
/db\_xref="taxon:10090"  
/clones="D230003K17"  
/tissue\_type="eyeball"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
473 - 2200  
/note="unnamed protein product; fucosyltransferase 8  
(MGD|MG1:1858901, GB|NM\_016893, evidence: BLASTN, 99%,  
match=2175)  
putative"  
/codon\_start=1  
/protein\_id="BAC34777.1"  
/db\_xref="GI:26342230"  
/translation="MRATGSGRWIMLIFAWGTLTLLFYIGHLVRDNDHPDHSRELS  
KILAKELRQKQNEELRMAESRIPEGLDQGTATGRVLEQLVKAQEIENIKK  
QARNLKGDEILRRRIENGAKELWFFLOSELKKLELQEGNLEQRLHDEILDLGHHE  
RSIMDLVYLSQTDGAGWREKAKDTELQVRITLYLQNPDKCKARKLVCNKGHC  
GYCOLHHVYVCFWIAVGTORTLILESONRWYATGWTVPVSETCTDRSGSTGH  
WSSEVNDKNQVVELPIVDSLHPPPLPLAVPEDLADRLRLVHGDPAVWVVSOFVKY  
LIRPQWLEKEIEATKGLGKGVIGVHVRRITDKVGTAAAFHPIEYMWVVERFQL  
LARMQVKKRVYLATDPTLLKEAKTKSYNEFISDINSWSAGLHNRITENSURVG  
ILDHIFLSQADFLVCTFSQVCRVAYEIMQTLHPDASANFHSLLDDIYFGQGNAHQI  
AVVPHKPTREEIPMEPGDIIGVAGNHWGVSKGINKRKLKGTGLYPSYKVREKIEYTK  
YPTYPEAK"

## CDS

2750..2755  
/note="putative"  
2766  
/note="putative"

## polyA\_signal

2750..2755

## polyA\_site

2766

BASE COUNT 838 a 591 c 619 g 718 t

## ORIGIN

Query Match 100.0%; Score 1728; DB 11; Length 2766;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCGGCATGACTGGTCTCTGGCGTGGATTATGCTCATTTCTTTTGGCTGGGGACC 60  
DB 473 ATGCGGCATGACTGGTCTCTGGCGTGGATTATGCTCATTTCTTTTGGCTGGGGACC 532  
QY 61 TTGTTATTTATATAGTGGTCTCATTTGTTTCGAGATAATGACCCCTGATCCTCCAGC 120  
DB 533 TTGTTATTTATATAGTGGTCTCATTTGTTTCGAGATAATGACCCCTGATCCTCCAGC 592  
QY 121 AGAGAACTCTCCAAGATTTCTTGAAGCTTGAACCGCTTAAACACGCAAAATGAAGACTTG 180  
DB 593 AGAGAACTCTCCAAGATTTCTTGAAGCTTGAACCGCTTAAACACGCAAAATGAAGACTTG 652  
QY 181 AGCGGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTACA 240  
DB 653 AGCGGAATGGCTGAGTCTCTCCGAATACCAAGAGCCCATTTGACCGGGGACAGCTACA 712  
QY 241 GGAAGAGTCCCTGTTTGTAGAGAAACAGCTTCTTAAAGCCCAAAGAACAGATTGAAATTTAC 300  
DB 713 GGAAGAGTCCCTGTTTGTAGAGAAACAGCTTCTTAAAGCCCAAAGAACAGATTGAAATTTAC 772  
QY 301 AAGAAAACAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAGAGAGAGGATTGAA 360  
DB 773 AAGAAAACAGCTAGAAATGGTCTGGGAAGGATCATGAAATCTTAAGAGAGAGGATTGAA 832  
QY 361 AATGAGCTAAGAGCTCTGTTTCTCAAAAGCGAAGCTGAAGAAATTTAAAGCATTTA 420  
DB 833 AATGAGCTAAGAGCTCTGTTTCTCAAAAGCGAAGCTGAAGAAATTTAAAGCATTTA 892  
QY 421 GAAGAAATGAATCCCAAGACATGCAGATGAAATCTTTTGGATTTTAGGACACCATGAA 480  
DB 893 GAAGAAATGAATCCCAAGACATGCAGATGAAATCTTTTGGATTTTAGGACACCATGAA 952  
QY 481 AGGCTATCATGACAGATCTTACTACTCTAGTCAACACAGATGGAGCAGGGGATTTGGCGT 540

Db 953 AGGTTATCATGACAGATCTACTACCTCAGTCAAAACAGATGGAGCAGGAGTTGGCGT 1012  
Qy 541 GAAAGAAGGCCAAGATCTGACGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 600  
Db 1013 GAAAGAAGGCCAAGATCTGACGAGCTGGTCCAGCGGAGAAATACATATCTCCAGAT 1072  
Qy 601 CCTAAGGACTGACGACAAAGCCAGGAAGCTGGTGTGTAAACATCAATATAAGGCTGGCTAT 660  
Db 1073 CCTAAGGACTGACGACAAAGCCAGGAAGCTGGTGTGTAAACATCAATATAAGGCTGGCTAT 1132  
Qy 661 GGTGTCACTCATCACTGGTGTACTGTGTTCATGATGTCTATATGGCAACCCAGCGAACA 720  
Db 1133 GGTGTCACTCATCACTGGTGTACTGTGTTCATGATGTCTATATGGCAACCCAGCGAACA 1192  
Qy 721 CTCATCTTGGAACTCTCAGAAATGGCGCTATGCTACTGCTGGATGGAGACTGTGTTAGA 780  
Db 1193 CTCATCTTGGAACTCTCAGAAATGGCGCTATGCTACTGCTGGATGGAGACTGTGTTAGA 1252  
Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGGTCCAGGTGAA 840  
Db 1253 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGGTCCAGGTGAA 1312  
Qy 841 GTAAATGACAAAACATTCAGAGTGTGAGCTGCCCTCCCATTTGTAGACAGCTCCATCCTCGG 900  
Db 1313 GTAAATGACAAAACATTCAGAGTGTGAGCTGCCCTCCCATTTGTAGACAGCTCCATCCTCGG 1372  
Qy 901 CCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCGACTCCTTAAGAGTCCAT 960  
Db 1373 CCTCCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCGACTCCTTAAGAGTCCAT 1432  
Qy 961 GGTGACCTCGAGTGTGGTGGTCCAGATTTGTCAAAATCTGATTCGTCACAACT 1020  
Db 1433 GGTGACCTCGAGTGTGGTGGTCCAGATTTGTCAAAATCTGATTCGTCACAACT 1492  
Qy 1021 TGGCTGGAAGAAATAGAGAACCCACCAAGAGCTTGGCTTCAAAACATCCAGTTAT 1080  
Db 1493 TGGCTGGAAGAAATAGAGAACCCACCAAGAGCTTGGCTTCAAAACATCCAGTTAT 1552  
Qy 1081 GGAGTCCATGTGACAGCGCACAGAAAGTGGAGACAGAGAGCTTCCACCCCATCGAG 1140  
Db 1553 GGAGTCCATGTGACAGCGCACAGAAAGTGGAGACAGAGAGCTTCCACCCCATCGAG 1612  
Qy 1141 GAGTACATGTTACAGCTTGAAGAACATTTTCAGCTTCTCGACGACAGAAATGCAAGTGGAT 1200  
Db 1613 GAGTACATGTTACAGCTTGAAGAACATTTTCAGCTTCTCGACGACAGAAATGCAAGTGGAT 1672  
Qy 1201 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTGTAAAGGAGGCAAGACAAAG 1260  
Db 1673 AAAAAAGAGTATATCTGGCTACTGATGATCCTACTTTTGTAAAGGAGGCAAGACAAAG 1732  
Qy 1261 TACTCCAATTATGAATTTATAGTGATAACTTATTTCTGGTCTGAGCTGGACTACAAAT 1320  
Db 1733 TACTCCAATTATGAATTTATAGTGATAACTTATTTCTGGTCTGAGCTGGACTACAAAT 1792  
Qy 1321 CGGTACACAGAAAATTCACCTCGGGGTGTGATCTCGGATATACACTTCTCTCACAGCT 1380  
Db 1793 CGGTACACAGAAAATTCACCTCGGGGTGTGATCTCGGATATACACTTCTCTCACAGCT 1852  
Qy 1381 GACTTTCTAGTGTGACTTTTTTCATCCAGGCTGTCTGGGTGCTTATGAATCATGCAA 1440  
Db 1853 GACTTTCTAGTGTGACTTTTTTCATCCAGGCTGTCTGGGTGCTTATGAATCATGCAA 1912  
Qy 1441 ACCCTGCATCTGATGCTCTGCGAATCTTCATTTCTTTGGATGATCATCTATTTTGGG 1500  
Db 1913 ACCCTGCATCTGATGCTCTGCGAATCTTCATTTCTTTGGATGATCATCTATTTTGGG 1972  
Qy 1501 GGCCTAAATGCCCAATCAGATTCGTTTATCTCTCAAACTCCGACCTGAGAGGAA 1560  
Db 1973 GGCCTAAATGCCCAATCAGATTCGTTTATCTCTCAAACTCCGACCTGAGAGGAA 2032  
Qy 1561 ATTCCAATGGAACTGGAGATATCATTTGGTGGCTGGAAACCATTTGGGATGGTTATCT 1620

Db 2033 ATTCCAATGGAACTGGAGATATCATTTGGTGTGGCTGGAAACCATTTGGATGGTTATCT 2092  
Qy 1621 AAAGGTATCAACAGAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 1680  
Db 2093 AAAGGTATCAACAGAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAGAG 2152  
Qy 1681 AAGTAGAAGACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATAG 1728  
Db 2153 AAGTAGAAGACAGTCAAGTATCCACATATCTCGAAGCTGAAAAATAG 2200

RESULT 2  
AK048520  
LOCUS  
DEFINITION  
Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone: C130068P22 product: fucosyltransferase 8, full insert sequence.  
AK048520  
VERSION  
AK048520.1 GI: 26339371  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1  
Carninci, P. and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE  
99279253  
PUBMED  
10349636

REFERENCE  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE  
20499374  
PUBMED  
11042159

REFERENCE  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL  
Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE  
20530913  
PUBMED  
11076861

REFERENCE  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Harada, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S. and Hayashizaki, Y.  
TITLE  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
Nature 409 (6821), 685-690 (2001)

MEDLINE PUBMED REFERENCE AUTHORS	21095660 11217851 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 3052) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, I., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kohji, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitho, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ Location/Qualifiers 1. .3052 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:CL30068P22" /db_xref="taxon:10090" /clone="Cl30068P22" /tissue_type="head" /clone_lib="RIKEN full-length enriched mouse cDNA library" 654. .2381 /note="unnamed protein product; fucosyltransferase 8 (MGD MG1:1858901, GB NM_016893, evidence: BLASTN, 99%, match=2175) putative" /codon_start=1 /protein_id="BAC33357.1" /db_xref="GI:26339372" /translation="MRWNTGSRWIMLILFAWGTLLFYIGHLVRDNDHPDHSRELS KILAKLERLKOQNEDLREMAESLRIPGPIDQGTATGVRVLEBOLVKAQEIENYKK QARNGLKDXEILRRRIENGAKELWFLQSELKKLHLEGNELORHADEILLDLGHE RSIMTDLTYLSDTDGADWRKEAKDLTELVRRIYQLQNPDCSKARKVCNINKGK GYGCQHVHYQVFIAYGTQRTLILESQNRWYATGGWETVFRPVSCTDRSLSTGH WSGEYNDKNIQVLEPIVDLSLHPPYPLFLAVPEDLADRLRHVGDPAVMWVSQFVKY LIRPOPWLEKIEEATKGLFKPHKPVIGVHVRRTKVGEAPHPFIEEYVHVEHFLV LARMQVDKRYLATDPTDLLKAKTKYSNVETSDNSISWSAGLHNRVYTESILRGV ILDIHFLSQADFVCTFFSSQVCRVAYEIMQTLHPDASANFHSLLDDIYYPFGQNAHQI AVYPHKPRTEIEIPMPGDIIGVAGNHWGDSKGINRKLKGTGLYPSYKVKREKLETVA YPTYPEAEK" 3036. .3041 /note="putative" 3052 /note="putative" polyA_signal polyA_site	Query Match Best Local Similarity 100.0%; Score 1728; DB 11; Length 3052; Matches 1728; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Qy 1 ATCGGGCGATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGGACC 60 Db 654 ATCGGGCGATGGACTGGTCTCTGGCGTTGGATTATGCTCATTTCTTTTTCCTGGGGGACC 713 Qy 61 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCACTGATCATCTCCAGC 120 Db 714 TTGTTATTTTATATAGTGGTTCATTTGGTTCGAGATAATGACCACTGATCATCTCCAGC 773 Qy 121 AGAGAATCTCTCAAGATTTCTTGAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180 Db 774 AGAGAATCTCTCAAGATTTCTTGAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 833 Qy 181 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCATTTGACCGGGGACAGCTACA 240 Db 834 AGGCGAATGGCTGAGTCTCTCCGAATACCAAGAGGCCCATTTGACCGGGGACAGCTACA 893 Qy 241 GGAAGAGTCCGTGTTTAGAAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAATTTAC 300 Db 894 GGAAGAGTCCGTGTTTAGAAGAAACAGCTTCTTAAGGCCAAAGAACAGATTGAAAATTTAC 953 Qy 301 AAGAAACAAAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAA 360 Db 954 AAGAAACAAAGCTAGAAATGGTCTCGGGAAGGATCATGAAATCTTTAAGAGGAGGATTGAA 1013 Qy 361 AATGAGCTAAGAGCTCTGGTTTTTCTCAAGCCGAACTGAAGAAATTAAGCATTTA 420 Db 1014 AATGAGCTAAGAGCTCTGGTTTTTCTCAAGCCGAACTGAAGAAATTAAGCATTTA 1073 Qy 421 GAAGCAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 480 Db 1074 GAAGCAATGAATCCAAAGACATGCAGATGAAATCTTTTGGATTTAGGACACCATGAA 1133 Qy 481 AGGTCTATCATGACAGATCTATCTACTCAGTCAAAACAGATGAGCAGGGGATTTGGCGT 540 Db 1134 AGGTCTATCATGACAGATCTATCTACTCAGTCAAAACAGATGAGCAGGGGATTTGGCGT 1193 Qy 541 GAAAGAGGCCAAGATCTGACAGAGCTGGTCAGCGGAGAAATTAACATATCTCCAGAT 600 Db 1194 GAAAGAGGCCAAGATCTGACAGAGCTGGTCAGCGGAGAAATTAACATATCTCCAGAT 1253 Qy 601 CCTAAGGACTGCACCAAGCCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 660 Db 1254 CCTAAGGACTGCACCAAGCCAGGAAGCTGGTGTGTAAACATCAATAAAGGCTGTGGCTAT 1313 Qy 661 GGTTGTCAACTCCATCACTGCTCTACTGTTTCATGATTTGTTAGCCACCCAGCAACA 720 Db 1314 GGTTGTCAACTCCATCACTGCTCTACTGTTTCATGATTTGTTAGCCACCCAGCAACA 1373 Qy 721 CTGATCTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGGAGACTGTGTTTGA 780 Db 1374 CTGATCTGGAATCTCAGAAATTTGGCGCTATGCTACTGTTGGATGGGAGACTGTGTTTGA 1433 Qy 781 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGCTCAGGTGAA 840 Db 1434 CCTGTAAGTGAGACATGTACAGACAGATCTGGCCCTCTCCACTGGACACTGCTCAGGTGAA 1493 Qy 841 GTAAATGACAAAAACATTCAGAGTGGTCCAGCTCCCAATTTGTAGACAGCTCCATCTCGG 900 Db 1494 GTAAATGACAAAAACATTCAGAGTGGTCCAGCTCCCAATTTGTAGACAGCTCCATCTCGG 1553 Qy 901 CCTCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCCGACTCTTAAGAGTCCAT 960 Db 1554 CCTCTTACTTACCACTGGCTGTTCCAGAGACCTTGCAGACCCGACTCTTAAGAGTCCAT 1613 Qy 961 GGTGACCTCGAGTGTGGTGGTGTCCAGTTTGTCAAATACTTGTTCGTCACCAACCT 1020 Db 1614 GGTGACCTCGAGTGTGGTGGTGTCCAGTTTGTCAAATACTTGTTCGTCACCAACCT 1673
---	--	---

BASE COUNT  
ORIGIN

1021 TGGCTGGAAGGAATAGAGAGCCACCAGAGCTTGGCTTCAACATCCAGTTATT 1080  
Db TGGCTGGAAGGAATAGAGAGCCACCAGAGCTTGGCTTCAACATCCAGTTATT 1733  
1081 GGAGTCCATGTCAGACGCACAGACAAAGTGGACAGAGCAGCTTCCACCCATCGAG 1140  
Db GGAGTCCATGTCAGACGCACAGACAAAGTGGACAGAGCAGCTTCCACCCATCGAG 1793  
1141 GAGTACATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAT 1200  
Db GAGTACATGGTACAGCTTGAAGAACATTTTCAGCTTCTCGCAGCAGAAATGCAAGTGGAT 1853  
1201 AAAAAAGAGATATATCGGCTACTGATGATCCTACTTTGTTAAAGAGGCAAGACAAAG 1260  
Db AAAAAAGAGATATATCGGCTACTGATGATCCTACTTTGTTAAAGAGGCAAGACAAAG 1913  
1261 TACTCCAATTATGAATTTATTAGTGAATACTTCTTGGTTCAGCTGGATACACAT 1320  
Db TACTCCAATTATGAATTTATTAGTGAATACTTCTTGGTTCAGCTGGATACACAT 1973  
1321 CGGTACACAGAAAATTCACCTCGGGGTGTGATCTCGGATATACACTTTCTCTCAGAGCT 1380  
Db CGGTACACAGAAAATTCACCTCGGGGTGTGATCTCGGATATACACTTTCTCTCAGAGCT 2033  
1381 GACTTTCTAGTGTGATCTTTTCATCCAGAGTCTGTGGGTGCTTATGAATCATGCAA 1440  
Db GACTTTCTAGTGTGATCTTTTCATCCAGAGTCTGTGGGTGCTTATGAATCATGCAA 2093  
1441 ACCCTGCATCTGATGCTCTGCGAATCTCAATCTTTGGATGATCACTATTTTGA 1500  
Db ACCCTGCATCTGATGCTCTGCGAATCTCAATCTTTGGATGATCACTATTTTGA 2153  
1501 GGCCAAAATGCCAATCAGATTGCTGTATCTCTCAAAACCTCGAACTGGAAGAGAA 1560  
Db GGCCAAAATGCCAATCAGATTGCTGTATCTCTCAAAACCTCGAACTGGAAGAGAA 2213  
1561 ATTCAAATGGAACCTGGAGATATCAATGGTGTGGTGGAAACCATTTGGATGGTTATCT 1620  
Db ATTCAAATGGAACCTGGAGATATCAATGGTGTGGTGGAAACCATTTGGATGGTTATCT 2273  
1621 AAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCAAAAGTCGAGAG 1680  
Db AAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCAAAAGTCGAGAG 2333  
1681 AAGATAGAACAGTCAAGTATCCCATATCTCTGAGCTGAAATAG 1728  
Db AAGATAGAACAGTCAAGTATCCCATATCTCTGAGCTGAAATAG 2381

RESULT 3  
LOCUS BQ892101  
DEFINITION BQ892101 937 bp mRNA linear EST 16-AUG-2002  
5', mRNA sequence.  
ACCESSION BQ892101  
VERSION BQ892101.1 GI:22284115  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 937)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Susan L. Sullivan, PhD.  
cDNA Library Preparation: ResGen, Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILLNL at:

<http://image.llnl.gov>  
Plate: L1AM13875 row: d column: 14  
High quality sequence start: 130  
High quality sequence stop: 780.  
FEATURES  
Location/Qualifiers  
1..937  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6389725"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_129"  
/note="Organ: olfactory epithelium; Vector:  
pCMV-SPO76.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned  
unidirectionally. Primer: Oligo dt. Average insert size  
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this  
is a NIH MGC Library." 244 t 1 others  
BASE COUNT 268 a 215 c 209 g 244 t  
ORIGIN  
Query Match 52.8%; Score 913.2; DB 13; Length 937;  
Best Local Similarity 99.5%; Pred. No. 4.5E-226;  
Matches 926; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
Qy 736 CAGAAATGGCGCTATGCTACTGTTGGATGGAGACTGTGTTAGACCTGTAAAGTGAACA 795  
Db 1 CAGAAATGGCGCTATGCTACTGTTGGATGGAGACTGTGTTAGACCTGTAAAGTGAACA 60  
Qy 796 TGTACAGACAGATCTGCGCTCTCCACTGGACACTGGTCAAGTGAATTAATGACAAAAAC 855  
Db 61 TGTACAGACAGATCTGCGCTCTCCACTGGACACTGGTCAAGTGAATTAATGACAAAAAC 120  
Qy 856 ATTCAAGTGGTTCAGAGTCCCGCATTTGTAGACAGCTCTCATCTCGGCTCTCTTACTTACCA 915  
Db 121 ATTCAAGTGGTTCAGAGTCCCGCATTTGTAGACAGCTCTCATCTCGGCTCTCTTACTTACCA 180  
Qy 916 CTGCTGCTTCAGAGAAACCTTTGCAGACCGACTCTCTAAGAGTCCATGTGAGCCCTGCAGTG 975  
Db 181 CTGCTGCTTCAGAGAAACCTTTGCAGACCGACTCTCTAAGAGTCCATGTGAGCCCTGCAGTG 240  
Qy 976 TGGTGGGTGCTCCAGTTGTCAAAATACCTTGTGTCGCAAAACCTTGGCTGGAAAAAGAA 1035  
Db 241 TGGTGGGTGCTCCAGTTGTCAAAATACCTTGTGTCGCAAAACCTTGGCTGGAAAAAGAA 300  
Qy 1036 ATAGAAGAGCCACCAAGAGCTTGGCTTCAAAACATCCAGTTATTGGAGTCCATGTGACA 1095  
Db 301 ATAGAAGAGCCACCAAGAGCTTGGCTTCAAAACATCCAGTTATTGGAGTCCATGTGACA 360  
Qy 1096 CGCACAGACAAAGTGAGAACAGAGCAGCTTCCACCCCATCGAGGAGTACATGGTACAC 1155  
Db 361 CGCACAGACAAAGTGAGAACAGAGCAGCTTCCACCCCATCGAGGAGTACATGGTACAC 420  
Qy 1156 GTTCAAGAACATTTTACGCTTCTCGACGCAAGATGCAAGTGGATATAAAGAGTATAT 1215  
Db 421 GTTCAAGAACATTTTACGCTTCTCGACGCAAGATGCAAGTGGATATAAAGAGTATAT 480  
Qy 1216 CTGCTACTGATGATCTACTTTGTTAAAGAGGCAAGAGAAAGTACTCCAATTATGAA 1275  
Db 481 CTGCTACTGATGATCTACTTTGTTAAAGAGGCAAGAGAAAGTACTCCAATTATGAA 540  
Qy 1276 TTTATTAGTGATACTCTATTCTTGTGCTGAGCTGAGCTACACAATCGGTACACAGAAAT 1335  
Db 541 TTTATTAGTGATACTCTATTCTTGTGCTGAGCTGAGCTACACAATCGGTACACAGAAAT 600  
Qy 1336 TCACCTCGGGGTGTGATCTGGATATACATTTCTCTCAGAGGTGAGCTTTCTAGTGTGT 1395  
Db 601 TCACCTCGGGGTGTGATCTGGATATACATTTCTCTCAGAGGTGAGCTTTCTAGTGTGT 660  
Qy 1396 ACTTTTCAATCCAGATCTGTGCGGTTGCTTTATGAATCATGCAACCTGCATCTCAT 1455  
Db 661 ACTTTTCAATCCAGATCTGTGCGGTTGCTTTATGAATCATGCAACCTGCATCTCAT 720  
Qy 1456 GCCTCTGCGAACTTCTATTCTTTTGGATGACATCTACTATTTTGGAGGCCAAATGCCCCAC 1515

Db 721 GGCCTCGGAACCTTCATCTTTGATGACATCTACTATTTTGGAGGCCAAATGCCAC 780  
QY 1516 AATCAGATTGCTGTTTATCTCACAACCTCGAACTGAAGAGAAATTCGAATGAACT 1575  
Db 781 AATCAGATTGCTGTTTATCTCACAACCTCGAACTGAAGAGAAATTCGAATGAACT 840  
QY 1576 GGAGATATCATTTGGTGGCTGGAACC-ATTGGATGGTTATTTCTAAAGGTATCAACAG 1634  
Db 841 GGAGATATCATTTGGTGGCTGGAACC-ATTGGATGGTTATTTCTAAAGGTATCAACAG 900  
QY 1635 AAAAAGTGGAAAAACAGGCTTATATCCCTCC 1665  
Db 901 AAAAAGTGGAAAAACAGGCTTATATCCCTTC 931

RESULT 4  
CB194828  
LOCUS  
DEFINITION  
AGENCY 11259261 NIH\_MGC\_135 Mus musculus cDNA clone  
IMAGE:30136221 5', mRNA sequence.  
ACCESSION  
CB194828  
VERSION  
CB194828.1 GI:28220043  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 898)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. David Rowe  
cDNA library preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM0035 row: h column: 22  
High quality sequence stop: 658.  
Location/Qualifiers  
1..898  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30136221"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_135"  
/note="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Normalized full-length enriched library from pooled mouse  
embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5,  
and 15.5 (size selected for the 0.5-1 kb fragments)  
Cloned directionally, priming method: Oligo-dT. cDNA  
enrichment: >1k bp. Average insert size 1.6k bp.  
Normalization (Cot value): 7.5 kb. Priming sequence:  
5'-GACTAGTTCTAGATCGGAGCGCGCCGCTTT-3' Tissue contributed by  
, David Rowe. Library constructed by Resgen, Invitrogen  
Corp."

## FEATURES

source  
BASE COUNT 272 a 203 c 193 g 230 t  
ORIGIN  
Query Match 46.4%; Score 801.6; DB 14; Length 898;  
Best Local Similarity 98.8%; Pred. No. 4.6e-197;  
Matches 818; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
QY 899 GGCCTCCTTACTTACCACTGGCTGTTCCAGAGAGACTTGCAGACCGACTCTCAAGAGTCC 958  
Db 1 GGCCTCCTTACTTACCACTGGCTGTTCCAGAGAGACTTGCAGACCGACTCTCAAGAGTCC 60  
QY 959 ATGGTGACCCCTGAGTGGTGGGTGCCAGTCTTTGTCAATFACCTTGATCGTCCACAC 1018

Db 61 ATGTGACCCCTGCGAGTGTGGGTGTCCTCCAGTTTGTCAATATCTGTTGTCACAC 120  
QY 1019 CTTGGCTGGAAAAGAAATAGAGAAGCCCAAGAAAGCTTTGGCTTCAACATCCAGTTA 1078  
Db 121 CTTGGCTGGAAAAGAAATAGAGAAGCCCAAGAAAGCTTTGGCTTCAACATCCAGTTA 180  
QY 1079 TTGAGTGCATGTGAGACGCAAGAAAGTGGGAAACAGAAAGAGCCCTTCCACCCCATCG 1138  
Db 181 TTGAGTGCATGTGAGACGCAAGAAAGTGGGAAACAGAAAGAGCCCTTCCACCCCATCG 240  
QY 1139 AGGAGTACATGTTGAGACGCAAGAAAGTGGGAAACAGAAAGAGCCCTTCCACCCCATCG 1198  
Db 241 AGGAGTACATGTTGAGACGCAAGAAAGTGGGAAACAGAAAGAGCCCTTCCACCCCATCG 300  
QY 1199 ATAAAAAAGAGTATATCTGGCTACTGATGATCTCTCTTTTAAAGAGGCAAGACAA 1258  
Db 301 ATAAAAAAGAGTATATCTGGCTACTGATGATCTCTCTTTTAAAGAGGCAAGACAA 360  
QY 1259 AGTACTCAATTTAGAAATTTATAGTGAATACTCTATTTCTTTGGTCACTGAGCTACACA 1318  
Db 361 AGTACTCAATTTAGAAATTTATAGTGAATACTCTATTTCTTTGGTCACTGAGCTACACA 420  
QY 1319 ATCGGTACACAGAAATTCACCTCGGGGTGTGATCTGATATACACTTTCTCTCACAGG 1378  
Db 421 ATCGGTACACAGAAATTCACCTCGGGGTGTGATCTGATATACACTTTCTCTCACAGG 480  
QY 1379 CTGACTTTCTAGTGTGACTTTTTTATCCCAAGGCTGTGCGGGTGTGCTTATGAATCATGC 1438  
Db 481 CTGACTTTCTAGTGTGACTTTTTTATCCCAAGGCTGTGCGGGTGTGCTTATGAATCATGC 540  
QY 1439 AAACCCCTGATCTGATGCTCTCGGAACCTTCCATTTCTTTGGATGACATCTACTATTTTG 1498  
Db 541 AAACCCCTGATCTGATGCTCTCGGAACCTTCCATTTCTTTGGATGACATCTACTATTTTG 600  
QY 1499 GAGGCCAAATGCCCCAATCAGATTCGTTTATCTCACAACCTCGAACTGGAAGAGG 1558  
Db 601 GAGGCCAAATGCCCCAATCAGATTCGTTTATCTCACAACCTCGAACTGGAAGAGG 660  
QY 1559 AAATTCCAATGGAACTCGGAGATATCATTTGGTGGCTGGAACCATTTGGATGTTATT 1618  
Db 661 AAATTCCAATGGAACTCGGAGATATCATTTGGTGGCTGGAACCATTTGGATGTTATT 720  
QY 1619 CTAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAG 1678  
Db 721 CTAAAGGTATCAACAGAAAACTTGGAAAAACAGGCTTATATCCCTCTCAAAAGTCCGAG 780  
QY 1679 AGAAGTAGAAGACAG-TCAGTATCCCATATCTCTGAAGCTGAAAAA 1725  
Db 781 AGAAGTAGAAGACAG-TCAGTATCCCATATCTCTGAAGCTGAAAAA 828

## RESULT 5

BM928521  
LOCUS  
DEFINITION  
AGENCY 6715254 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:5798262  
5', mRNA sequence.  
ACCESSION  
BM928521  
VERSION  
BM928521.1 GI:19378900  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1066)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)



DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: L1CM2025 row: p column: 07  
 High quality sequence stop: 666.

# FEATURES

Location/Qualifiers  
 1. .1066  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5798262"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_100"

Note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 315 a 235 c 248 g 264 t 4 others

## ORIGIN

Query Match 45.6%; Score 787.6; DB 12; Length 1066;  
 Best Local Similarity 89.2%; Pred. No. 2.1e-193;  
 Matches 880; Conservative 0; Mismatches 103; Indels 3; Gaps 3;

QY 366 AGCTAAAGAGCTCTGGTCTTTCTCAAGCGCACTGAAGAAATTAAGCACTTTAGAAGG 425  
 DB 1 AGCTAAAGAGCTCTGGTCTTTCTCAAGCGCACTGAAGAAATTAAGCACTTTAGAAGG 60  
 QY 426 AAATGAATCCAAAGACATGCAGATGAATTTCTTTGGATTGAGACACCATGAAGGTC 485  
 DB 61 AAATGAATCCAAAGACATGCAGATGAATTTCTTTGGATTGAGACATCATGAAGGTC 120  
 QY 486 TATCATGACAGATCTATCTACTCTCAGTCAACACATGAGCGGGATTGGCGTGAAGA 545  
 DB 121 TATAATGACGGATCTATCTACTCTCAGTCAACACATGAGCGGGATTGGCGGGA 180  
 QY 546 AGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCTAA 605  
 DB 181 AGAGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCTAA 240  
 QY 606 GGACTGCGAGCAAGCCAGGAAGCTGGTGTGAACATCAATAAAGCTGTGGCTATGGTTG 665  
 DB 241 GGACTGCGAGCAAGCCAAAGCTGGTGTGAATAATCAACAAAGCTGTGGCTATGGCTG 300  
 QY 666 TCAACTCCATCAGTGGTCTACTGTTCATGATTCCTTATGACCCAGCGAGACATCAT 725  
 DB 301 TCAGCTCCATCATGTGGTCTACTGTTCATGATTCCTTATGACCCAGCGAGACATCAT 360  
 QY 726 CTTGGAATCTCAGATTTGGCGTCTACTGTGGTGGGAGACTGTGTTAGACCTGT 785  
 DB 361 CTTGGAATCTCAGATTTGGCGTCTACTGTGGTGGGAGACTGTGTTAGACCTGT 420  
 QY 786 AAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTTCAGGTGAAGTAA 845  
 DB 421 AAGTGAGACATGTACAGACAGATCTGGCTCTCCACTGGACACTGGTTCAGGTGAAGTAA 480  
 QY 846 TGACAAAACATTTAAGTGGTTCAGCTCCCATTTGATGACAGCTCCATCTCGGCTCC 905  
 DB 481 GGACAAAATGTTCAAGTGGTTCAGCTCCCATTTGATGACAGCTCCATCTCGGCTCC 540  
 QY 906 TTACTTACCACTGGCTGTTCAGAGAGACTTGCACAGCTCCCTAAGAGTCCATGGA 965  
 DB 541 ATATTACCTTGGCTGTACCAAGAGACTTCGCATCGACTTGTACAGGTGATGGGA 600  
 QY 966 CCCTGCAGTGTGGTGGGTGCCAGTTGTTCAATACCTTTGATTCGTCCACAACTTGGCT 1025

DB 601 CCCTGCAGTGTGGTGGTGTCTCTCAGTTGTCAAAATCTGATCCGCCACAGCCCTTGGCT 660  
 QY 1026 GGAAAGGAAATAGAAAGAGCCACCAGAAAGCTTGGCTTTCAAAATCATCAGTTATTGGAGT 1085  
 DB 661 AGAAAGGAAATAGAAAGAGCCACCAGAAAGCTTGGCTTTCAAAATCATCAGTTATTGGAGT 720  
 QY 1086 CCATGTGACGCGACAGACAAAGTGGGAAACAGAGACAGCTTCCACCCCATCGAGAGTA 1145  
 DB 721 CCATGTGACGCGACAGACAAAGTGGGAAACAGAGAGCTGCCCTTCCATCCCATTAAGAGTA 780  
 QY 1146 CATGGTACACCTTGAAGGACATTTTCAGCTTCTCGACGACAGATGCAAGTGGATGATAAAA 1205  
 DB 781 CATGGTACATGTTGAAGGACATTTTCAGCTTCTCGACGACAGATGCAAGTGGATGATAAAA 840  
 QY 1206 AAGAGTATATCTGGTCTACTGATGATCTCTTCTTTAAAGGAGGCAAGACAAAGTACTC 1265  
 DB 841 AAGAGTATATCTGGTCTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 1266 CAATTATGAATTT-ATTAGTGAATCACTTATTT-CTTGGTCACTGGAGTACAACTCGG 1323  
 DB 901 CAATTATGAATTTAATAATGATAACTCTATTTCCTGTCAGCGGAGCTTGGCCCATCGN 960  
 QY 1324 -TACACAGAAATTCACCTTCGGGGTG 1348  
 DB 961 ATACCCCGAAAAATTCCTCCGGGG 986

## RESULT 6

LOCUS BQ653329 940 bp mRNA linear EST 15-JUL-2002  
 DEFINITION AGENCOURT 8298428 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6270458  
 5', mRNA sequence.  
 ACCESSION BQ653329  
 VERSION BQ653329.1 GI:21777501  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 940)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1CM2445 row: k column: 03  
 High quality sequence stop: 700.

## FEATURES

Location/Qualifiers  
 1. .940  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6270458"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_100"  
 Note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 286 a 197 c 222 g 235 t

```
ORIGIN
Query Match 43.2%; Score 747.2; DB 13; Length 940;
Best Local Similarity 90.6%; Pred. No. 6.5e-183;
Matches 852; Conservative 0; Mismatches 83; Indels 5; Gaps 5;

QY 366 AGCTAAAGAGCTCTGGTGTCTTTTCTACAAAGCAAGCTGAAGAATTAAGCATTTAGAAGG 425
DB 1 AGCTAAAGAGCTCTGGTGTCTTTTCTACAGAGTGAATTTGAAGAAATTAAGAACTTTAGAAGG 60
QY 426 AAATGAACCTCCAAAGACATGACAGATCAATCTTTTGGATTATAGACACCATGAAGGTC 485
DB 61 AAATGAACCTCCAAAGACATGACAGATCAATCTTTTGGATTATAGACATCATGAAGGTC 120
QY 486 TATCATGACAGATCTATACCTCAGTCAACACAGATGGAGCAGGGGATTGGCGTGAAGA 545
DB 121 TATAATGACGGATCTATACCTCAGTCAGACAGATGAGCAGGTGATTGGCGGGAAGA 180
QY 546 AGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATAACATATCTCCAGAACTCTAA 605
DB 181 AGAGGCCAAAGATCTGACAGAACTGGTTCAGCGGAGATAACATATCTTCAGAACTCCAA 240
QY 606 GCAGTCGACAAAGCCAGAAAGCTGGTGTAAACATCAATAAAGCTGTGGCTATGGTTG 665
DB 241 GGACTGACAGAAAGCCAAAGCTGGTGTAAATATCAACAAAGCTGTGGCTATGGTTG 300
QY 666 TCAACTCCATCACTGCTGCTACTGTTTTCATGATTGCTTATGCGACCCAGCGAACAATCAT 725
DB 301 TCAGCTCCATCATGTTGGTCTACTGCTTCATGATTGCATATGCGACCGAACAATCAT 360
QY 726 CTTGGAATCTGAATTTGGCGTATGCTACTGGTGGATGGGAGACTGTGTTAGACCTGT 785
DB 361 CTTGGAATCTGAATTTGGCGTATGCTACTGGTGGATGGGAGACTGTGTTAGGCGCTGT 420
QY 786 RAGTGAGATCTGACAGACATCTGGCTCTCCACTGGACACTGGTCAAGTGAAGTAA 845
DB 421 AAGTGAGATCTGACAGACATCTGGCATCTCCACTGGACACTGGTCAAGTGAAGTAA 480
QY 846 TGACAAAAACATTTCAAGTGGTGGAGTCCCAATTTGACAGACCTTCCATCTCGCGCTCC 905
DB 481 GGACAAAATGTTCAAGTGGTGGAGTCCCAATTTGACAGACTCTTCATCCCGTCCCTCC 540
QY 906 TTACTTTACACCTGGTGTTCAGAGACACTTGGACAGCAGCTCTTAAAGATCCATGGTGA 965
DB 541 ATATTTTACCCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCATGGTGA 600
QY 966 CCTCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1025
DB 601 CCTCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 660
QY 1026 GGAAAGAGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTATTGGAGT 1085
DB 661 AGAAAAGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAAACTCCAGTTATTGGAGT 719
QY 1086 CCATGTTCAGACGACAGACAAAGTGGGAAACAGAGAGCTTTCACCCCATCGAGAGTGA 1145
DB 720 CCATGTTCAGACGACAGACAAAGTGGGAAACAGAGAGCTTTCACCCCATCGAGAGTGA 779
QY 1146 CATGGT-ACACGTTGAGAGCAATTTTCAGCTTCTCGCAGCAGATGCAAGTGGATAAA 1204
DB 780 CATGGTGGCATGTTGAAGAACATTTTCAGCTTCTTCGACGCAAAATGCAAGTGGACCAA 839
QY 1205 AAAGAGTATA-TCTGGCTACTGATCTCTCTCTTGT-TAAAGAGGCAAAAGACAAAGTA 1262
DB 840 AAAGAGTGTATTTTGGCCACAGATGACCTTCTTTTATAGAGGAGCAAAACCAAGGA 899
QY 1263 CTCCTAAATTAGAA-TTTATAGTGAATACTTATTTCTTG 1301
DB 900 CCCCCATATGAATTTTATATGGATAACTTATTTCTCTG 939

FEATURES
source
1..900
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS00DG005Y007"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT 250 a 201 c 200 g 245 t 4 others
ORIGIN
Query Match 43.2%; Score 746.8; DB 13; Length 900;
Best Local Similarity 90.3%; Pred. No. 8.1e-183;
Matches 807; Conservative 0; Mismatches 86; Indels 1; Gaps 1;

QY 633 GTGTAAACATCAATAAAGCTGTGGCTATGGTGTGCAACTCCATCAGCGTGTCTACTGTTT 692
DB 5 GTGTGTATATAACAAAGCTGTGGCTATGGCTGTGAGTCCAT-ATGGGTCTACTGCTT 63
QY 693 CATGATTGCTTTATGGCACCAGCGAACACATCATCTTGGAAATCTCAGAAATTCGCGCTATGC 752
DB 64 CATGATTGCATATGGCACCAGCGAACACATCATCTTGGAAATCTCAGAAATTCGCGCTATGC 123
QY 753 TACTGGTGGATGGGAGACTGTGTTTAGACCTGTGAAGTGAACATGTACAGACATCTGG 812
DB 124 TACTGGTGGATGGGAGACTGTATTTAGGCTGTGAAGTGAACATGTACAGACATCTGG 183
QY 813 CCTCTCCACTGGACACTGGTCCAGTGAAGTAAATGACAAAACATTCAGTGGTTCGAGCT 872
DB 184 CATCTCCACTGGACACTGGTCCAGTGAAGTGAAGCAAAATGTTCAAGTGGTTCGAGCT 243
QY 873 CCCCATTTGTAGACAGCTCCATCTCGGCTCTCTTACTTACCACTGGCTGTTCAGAGA 932
DB 244 TCCCATTTGTAGACAGCTCTTCATCCCGCTCTCCATATTTACCCCTTGGCTGTACCAAGA 303
QY 933 CCTTGCAGACCGACTCCTTAAGAGTCCATGTGACCCCTGCAGTGTGGTGGTGTCCAGTT 992
DB 304 CCTCGCAGATCGACTTGTGACAGTGTGGTGTGACCCCTGCAGTGTGGTGGTGTCTCAGTT 363
QY 993 TGTCAAAATCTTGAATTCGTCCACAACCTTGGCTCGAAAAGGAAATAGAAAGGCCACCA 1052
```

Db 364 TGTCAATPACTTGATCCGCCACACGCTTGCTAGAAAAAGAAATAGAAAGCCACCA 423  
Qy 1053 GAAGCTTGGCTTCAACATCAGTATTTGGAGTCCATGTACAGCGCAGACAAAGTGGG 1112  
Db 424 GAAGCTTGGCTTCAACATCAGTATTTGGAGTCCATGTACAGCGCAGACAAAGTGGG 483  
Qy 1113 AACAGAAGCAGCTTCCACCCATCGAGGAGTACATGTACACGTTGAAGAACATTTTCA 1172  
Db 484 AACAGAAGCAGCTTCCACCCATCGAGGAGTACATGTGTGATGTTGAAGAACATTTTCA 543  
Qy 1173 GCTTCTCGCAGCAGATGCAAGTGGATAAAAAAGAGTATATCTGGCTACTGATGCC 1232  
Db 544 GCTTCTCGCAGCAGATGCAAGTGGATAAAAAAGAGTATATTTGGCCACAGATGACCC 603  
Qy 1233 TACTTTGTTAAAGAGGAGGCAAGAGTACTTCCAAATTTATGATGAATTTATAGTGATACTC 1292  
Db 604 TCTTTTAAAGAGGAGGCAAGAGTACTTCCAAATTTATGATGAATTTATAGTGATACTC 663  
Qy 1293 TATTTCTTGGTCACTGGAGTACCAATTCGGTACACAGAAATTCACCTTCGGGGTGTGAT 1352  
Db 664 TATTTCTTGGTCACTGGAGTACCAATTCGGTACACAGAAATTCACCTTCGGGGTGTGAT 723  
Qy 1353 CCTGATATACATTTCTCTCACAGGCTGACTTTCTAGTGATGTTCTTCTATCCAGGT 1412  
Db 724 CCTGATATACATTTCTCTCACAGGCTGACTTTCTAGTGATGTTCTTCTATCCAGGT 783  
Qy 1413 CTGTCGGGTGCTTATGAATCATGCAAAACCTGCTGATGCTCTGCGAACTTCCA 1472  
Db 784 CTGCGAGTGTCTTATGAATTTATGCAAACTACATCTGATGCTCTGCGAACTTCCA 843  
Qy 1473 TTTCTTGGATGACATCTACTATTTTGGAGGCAAAATCCCAATCAGATGTC 1526  
Db 844 TTTCTTGGATGACATCTACTATTTTGGAGGCAAAATCCCAATCAGATGTC 897

RESULT 8  
BO423860  
LOCUS  
DEFINITION AGENCOURT\_7894821 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6158200  
5', mRNA sequence.  
ACCESSION BO423860  
VERSION BO423860.1 GI:21119175  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 886)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILLNL at:  
<http://image.illnl.gov>  
plate: LLAM13505 row: e column: 17  
High quality sequence stop: 641.  
Location/Qualifiers  
1..886  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6158200"  
/issue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 72"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
FEATURES  
source

Average insert size 2 kb. Library constructed by Life Technologies." BASE COUNT 273 a 185 c 212 g 215 t 1 others  
ORIGIN  
Query Match 42.5%; Score 735.2; DB 13; Length 886;  
Best Local Similarity 90.7%; Pred. No. 8.4e-180;  
Matches 794; Conservative 0; Mismatches 79; Indels 2; Gaps 1;  
Qy 408 ATTAAGACATTTAGAGGAATGAATCCAAAGACATGCAGATCAAAATCTCTTTGGATTT 467  
Db 1 ATTAAGACATTTAGAGGAATGAATCCAAAGACATGCAGATCAAAATCTCTTTGGATTT 60  
Qy 468 AGGACACCATGAAGTCTTATCATGACAGATCTATACCTCAGTCAACAGATGGAGC 527  
Db 61 AGGACATCATGAAGTCTTATATGACGGATCTATACCTCAGTCAACAGATGGAGC 120  
Qy 528 AGGGGATGGCGTGAAGAGGCAAGATCTCAGAGAGTGTGCCAGCGGAATAAC 587  
Db 121 AGGTGATTTGGCGGAAAGAGGCAAGATCTCAGAGACTGGTTTCAGCGGAATAAC 180  
Qy 588 ATATCTCCAGAACTCCYAAGGACTGCAGCAAGCCAGGAAGCTGTGTGTAACATCAATA 647  
Db 181 ATATCTTCAGAACTCCYAAGGACTGCAGCAAGCCAGGAAGCTGTGTGTAATAATCAACA 240  
Qy 648 AGGCTGTGGCTATGGTGTCAAACTCCATCCAGTGGTCTTCTATGATTTGCTTATGG 707  
Db 241 AGGCTGTGGCTATGGTGTCAAACTCCATCCATCCAGTGGTCTTCTATGATTTGCTATGG 300  
Qy 708 CACCAGCGAACTATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGTGATGGGA 767  
Db 301 CACCAGCGAACTATCTTTGGAATCTCAGAAATGGCGCTATGCTACTGTGATGGGA 360  
Qy 768 GACTGTCTTTAGACCTGTAAGTGAGACATGCACAGACAGATCTGCGCTCTCCACTGGACA 827  
Db 361 GACTGTATTTAGGCTGTAAGTGAGACATGCACAGACAGATCTGCGCTCTCCACTGGACA 420  
Qy 828 CTGCTCAGGTGAAGTAAATGACAAAAACATTCAGATGGTCCAGCTCCCAATTTAGACAG 887  
Db 421 CTGCTCAGGTGAAGTAAAGGACAAATGTTCAAGTGGTCCAGCTCCCAATTTAGACAG 480  
Qy 888 CCTCCATCTCGGCTCTCTTACCTTACCACTGGCTGTTCAGAGACCTTCGAGACCGACT 947  
Db 481 TCTTCATCCCGCTCTCTCATATTTACCTTGGCTGTACACAGACAGCTTCGAGATCGACT 540  
Qy 948 CTTAAGAGTCCATGCTGACCTGCACTGTGCTGGTGTGCCAGTTCCTCAATCTTGCAT 1007  
Db 541 TGTACGAGTGCATGGTGAACCTGCACTGTGCTGGTGTGCTCAGTTGTCAATACTTGCAT 600  
Qy 1008 TCGTCCACAACTTTGGCTGGAAAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAA 1067  
Db 601 CCGCCACAGCTTGGCTAGAAAAAGAAATAGAAAGCCACCAAGAGCTTGGCTTCAA 660  
Qy 1068 ACATCCAGTTATTTGGAGTCCATGTACAGCGCAGACAAAGTGGGAACAGAGCGCTT 1127  
Db 661 ACATCCAGTTATTTGGAGTCCATGTACAGCGCAGACAAAGTGGGAACAGAGCGCTT 720  
Qy 1128 CCACCCATCGAGAGTACATGGTACAGTTGAAGAACATTTTCAGCTTCTCGCAGCGAG 1187  
Db 721 CCATCCCATTTGAAGAGTACATGGTGCATGTTTGAAGAACATTTTCAGCTTCTCGCAGCGAG 780  
Qy 1188 AATGCAAGTGGATAAAAAAGAGTATATCTGGCTACTGATGATCC--TACTTTTGTAAAG 1245  
Db 781 AATGCAAGTGGACAAAAAGAGTGTATTTGGCCACAGATGACCCCTCTTTTAAAGG 840  
Qy 1246 GAGGCAAGACAAAGTACTCCAAATTTATCAATTTAT 1280  
Db 841 GAGGCAAGACAAAGTACTCCCAATTTATGAANTTTT 875

RESULT 9  
BX429921  
LOCUS  
BX429921  
923 bp mRNA linear EST 15-MAY-2003

DEFINITION BX429921 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
Clone CS0DG005Y007 5-PRIME, mRNA sequence.  
ACCESSION BX429921  
VERSION BX429921.1 GI:30780877  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 923)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7499.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAA012ZA08 CS01114.1&cluster=7499.f.  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0BAA012ZA08\_CS01114.1.  
Location/Qualifiers  
1..923  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DG005Y007"  
/tissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 285 a 184 c 231 g 218 t 5 others  
ORIGIN  
Query Match 42.0%; Score 725; DB 13; Length 923;  
Best Local Similarity 91.3%; Pred. No. 3.8e-177;  
Matches 811; Conservative 0; Mismatches 73; Indels 4; Gaps 4;  
QY 324 GGGGAAGGATCATG-AAATCTTTAAGAGGAGGATTGAAATGGAGCTAAAGAGCTCTGGT 382  
DB 31 GGGNAGNATCATGNAATCTCTGAGGAGGAGATTGAAATGGAGCTAAAGAGCTCTGGT 90  
QY 383 TTTTCTCAAGCGAAGCTGGAAGAAATTAAGCATTTTGAAGGAATGAATCTCAAGAC 442  
DB 91 TTTTCTCAAGAGTGAATTTGAAGAAATTAAGAACTTAGAAGAAATGAATCTCAAGAC 150  
QY 443 ATGAGATGAATTTCTTTGGATTAGACACCATGAAGGTCTATCATGACAGATCTAT 502  
DB 151 ATGAGATGAATTTCTTTGGATTAGACATCATGAAGGTCTATAATGACGGATCTAT 210  
QY 503 ACTACTCAGTCAACACAGATGAGCAGGGGATTGGCGTGAAAGAGCCAAAGATCTGA 562  
DB 211 ACTACTCAGTCAACAGATGAGCAGGGTATTGGCGGAAAGAGCCAAAGATCTGA 270  
QY 563 CAGAGCTGTCCAGCGGGAATTAACATATCTCAGAATCTTAAGAGCTGCAGCAAGCCA 622  
DB 271 CAGAAGCTGTTCAGCGGGAATTAACATATCTCAGAATCTTAAGAGCTGCAGCAAGCCA 330  
QY 623 GGAAGCTGTGTGAACATCAATTAAGGCTGTGGCTATGGTGTCACTCCATCATGCTGG 682  
DB 331 AAAAGCTGTGTGAATTAACAAAGGCTGTGGCTATGGCTGTGAGCTCCATCATGTGG 390  
QY 683 TCTACTGTTTCATGATGCTTATGACCCAGCAACTCATCTTGAATCTCAGAAATT 742  
DB 391 TCTACTGTTTCATGATGCTTATGACCCAGCAACTCATCTTGAATCTCAGAAATT 450

QY 743 GGCCTATGCTACTGCTGGATGGAGACTGTGTTTAGACCTGTAAAGTGAACATGTACAG 802  
DB 451 GGCCTATGCTACTGCTGGATGGAGACTGTGTTTAGACCTGTAAAGTGAACATGTACAG 510  
QY 803 ACAGATCTGGCTCTCCACTGGACACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 862  
DB 511 ACAGATCTGGCTCTCCACTGGACACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 570  
QY 863 TGGTCGAGCTCCCATTTGTAGACAGCTCCATCTCGGCTCTCTTACTTACCATCTGCTG 922  
DB 571 TGGTCGAGCTCCCATTTGTAGACAGCTCTCTCTCCGCTCTCTCTCTCTCTCTCTCTG 630  
QY 923 TTCAGAAAGACCTTGCAGACCGACTCTTAAGAGTCCATGGTGACCTGCAGTGTGGTGG 982  
DB 631 TACCAGAAAGACCTGCAGATCGACTTGTACAGTGGTGGTGGTGGTGGTGGTGGTGG 690  
QY 983 TGTCCAGTTTGTCAAAATACTTGAATGCTGCACAACTTGGCTGGAAGAAATAGAAAG 1042  
DB 691 TGTCTCAGTTTGTCAAAATACTTGAATGCTGCACAACTTGGCTGGAAGAAAG-ATAAG 749  
QY 1043 AAGCCACCAAGAAGCTTGGCTTCAAAATCCAGTTATTGGAGTCCATGTCAGACGACAG 1102  
DB 750 AAGCCACCAAGAAGCTTGGCTTCAAAATCCAGTTATTGGAGTCCATGTCAGACGACAG 808  
QY 1103 ACAAGTGGGACAGAGCAGCTTCCACCCATCGAGGAGTACATGTTGATCA-CTTTGAA 1161  
DB 809 ACAAGTGGGACAGAGCAGCTTCCATCCATTTGAAGAGTACATGTTGATGTTGAA 868  
QY 1162 GAACATTTCAGCTTCTCGACGAGCAATGCAAGTGGATGATGATGATGATGATGATG 1209  
DB 869 GAACATTTCAGCTTCTCGACGAGCAATGCAAGTGGATGATGATGATGATGATGATG 916  
RESULT 10  
LOCUS BU184296  
DEFINITION AGENCOURT\_7971158 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6171012  
5', mRNA sequence.  
ACCESSION BU184296  
VERSION BU184296.1 GI:22698280  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 855)  
NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13538 row: k column: 13  
High quality sequence stop: 711.  
Location/Qualifiers  
1..855  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6171012"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life

```
Technologies."
BASE COUNT      265 a 172 c 224 g 193 t 1 others
ORIGIN

Query Match      41.7%; Score 720.4; DB 13; Length 855;
Best Local Similarity 90.9%; Pred. No. 5.8e-176;
Matches 766; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 140 TTGCAAGCTTGAACGCTTAAACAGCAAGCAATGAAGACTTGAGGCGCAATGCTGAGTCTC 199
Db 12 TGGCAAGCTTGAAGCTTAAACACAGATGAAGACTTGAGGCGATGCCGATCTC 71

Qy 200 TCCGAATACCAAGGCCCTTAACAGAGGAGAGCTACAGGAGAGATCCGCTGTTTAC 259
Db 72 TCCGATACCAAGGCCCTTAACAGAGGAGAGCTATAGGAGAGATAGGAGAGTACGCGTTTAC 131

Qy 260 AAGACAGCTTGTAAAGCCCAAGACAGATGAAGATTAAGAAACAACTAGAAATG 319
Db 132 AAGAGCAGCTTGTAAAGCCCAAGACAGATGAAGATTAAGAAACAACTAGAAATG 191

Qy 320 GTCTGGGGAAGGATCATGAATCTTAAGAGGAGGATTAAGAAATGGAGCTAAAGAGCTCT 379
Db 192 GTCTGGGGAAGGATCATGAATCTTGAGAGGAGGATTAAGAAATGGAGCTAAAGAGCTCT 251

Qy 380 GGTCTTTTCTACAAAGCGAATGAAGAAATTAAGCAATTAAGAGAAATGAATCTCAAA 439
Db 252 GGTCTTTTCTACAGAGTGAATTAAGAAATTAAGAACTTAAGAGAAATGAATCTCAAA 311

Qy 440 GACATGAGATGAATCTTTTGGATTTAGACACCATGAAGGTCTATCATGACAGATC 499
Db 312 GACATGAGATGAATCTTTTGGATTTAGGACATCATGAAGGTCTATATATGACGGATC 371

Qy 500 TATACCTACTCAGTCAAAACAGATGAGCAGGGATTTGGCGTGAAGAGCCCAAGATC 559
Db 372 TATACCTACTCAGTCAAAACAGATGAGCAGGGATTTGGCGTGAAGAGCCCAAGATC 431

Qy 560 TGACAGAGCTGGTCCAGCGAGAAATCAATATCTCAGAAATCTTAAGAGCTGCAGCAAG 619
Db 432 TGACAGAACTGGTTCAGCGAGAAATCAATATCTTCAAGATCCCAAGAGCTGCAGCAAG 491

Qy 620 CCAGAAAGCTGGTGTAAACATCAATTAAGAGCTGGGTATGGTGTCAACTCCATCAG 679
Db 492 CCAAAAGCTGGTGTAAATATCAAAAGGCTGTGGCTATGGCTGTGAGCTCCATCATG 551

Qy 680 TGGTCTACTGTTCATGATGCTTATGCGACCCAGCAACATCATCTTGAATCTCAGA 739
Db 552 TGGTCTACTGTTCATGATGCTTATGCGACCCAGCAACATCATCTTGAATCTCAGA 611

Qy 740 ATTGGCGCTATGCTACTGGTGGAGAGACTGTGTTTGAACCTGTAAAGTGAAGATGTA 799
Db 612 ATTGGCGCTATGCTACTGGTGGAGAGACTGTATTTAGGCGCTGAAGTGAAGATGCA 671

Qy 800 CAGACAGATCTGGCTCTCCACTGGACACTGGTCAAGGTGAAGTAAATGACAAAACATTC 859
Db 672 AAGACAGATCTGGCATCTCCACTGGACACTGGTCAAGGTGAAGTGAAGACAAAATGTC 731

Qy 860 AAGTGGTCAAGTCCCAATTTAGACAGCTCCATCTCGGCTCTTACTTACCCTGG 919
Db 732 AAGTGGTCAAGTCCCAATTTAGACAGCTCTTACCCCGCTCCCTCATATTTACCCCTGG 791

Qy 920 CTGTTCCAGAAAGACCTTCAGACAGCTCCCTAAAGAGTCCATGGTGGACCTGCAGTGGT 979
Db 792 CTGTACAGAAAGACCTTCAGATCGACTTGTACGAGTGCATGGTGACCTGGCCNGGTGG 851

Qy 980 GGG 982
Db 852 GGG 854

RESULT 11
BU513162
LOCUS
DEFINITION AGENCOURT_10116686 NIH_MGC_134 Mus musculus cdna clone

IMAGE:6508154 5', mRNA sequence.
BU513162
VERSION
BU513162.1 GI:22820688
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 937)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrived by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14073 Row: g Column: 03
High quality sequence stop: 751.
FEATURES
Location/Qualifiers
1..937
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6508154"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/notes="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT      260 a 220 c 204 g 250 t 3 others
ORIGIN

Query Match      40.8%; Score 704.6; DB 13; Length 937;
Best Local Similarity 91.4%; Pred. No. 7.8e-172;
Matches 819; Conservative 0; Mismatches 5; Indels 72; Gaps 4;

Qy 690 TTTCATCATTTGCTTANGGCCACCCAGCAACTCATCTTGGAACTCTCAGAAATGGCGCTA 749
Db 1 TTTCATCATTTGCTTANGGCCACCCAGCAACTCATCTTGGAACTCTCAGAAATGGCGCTA 60

Qy 750 TGCTACTGGTGGATGAGAGACTGTGTTTAGACCTGTAAAGTGAGACATGTACAGACAGATC 809
Db 61 TGCTACTGGTGGATGAGAGACTGTGTTTAGACCTGTAAAGTGAGACATGTACAGACAGATC 120

Qy 810 TGGCTCTCCACTGGACACTGGTTCAGTGAAGTAAATGACAAAACATTCAGTGGTCTGA 869
Db 121 TGGCTCTCCACTGGACACTGGTTCAGTGAAGTAAATGACAAAACATTCAGTGGTCTGA 180

Qy 870 GCTCCCATTTGTAGACAGCTCCATCTCGSCCTCTTACTTACCATGGCTGTTCCAGA 929
Db 181 GCTCCCATTTGTAGACAGCTCCATCTCGSCCTCTTACTTACCATGGCTGTTCCAGA 240

Qy 930 AGACCTTTCGACAGCCGACTCCCTAAGAGTCCATGTGTGACCTGACCTGAGTGGTGGTTC 989
Db 241 AGACCTTTCGACAGCCGACTCCCTAAGAGTCCATGTGTGACCTGACCTGAGTGGTGGTTC 300

Qy 990 GTTTGTCAATATCTTATTTGTCACAACTTCAGTGAAGAAAGAAATAGAAAGCCAC 1049
Db 301 GTTTGTCAATATCTTATTTGTCACAACTTCAGTGAAGAAAGAAATAGAAAGCCAC 360

Qy 1050 CAAGAAGCTTGGCTTCAACATCCAGTATTTGGAGTCCATGTGTGAGCGCAGACACAAGT 1109
Db 361 CAAGAAGCTTGGCTTCAACATCCAGTATTTGGAGTCCATGTGTGAGCGCAGACACAAGT 420

Qy 1110 GGGAAACAGAGCAGCTTTCCACCCCATCGAGGAGTACATGGTACACGTTGAAGAACATTT 1169
Db 1110 GGGAAACAGAGCAGCTTTCCACCCCATCGAGGAGTACATGGTACACGTTGAAGAACATTT 1169
```

Db 421 GGAACAGAGAGAGCGCTTCCACCCATCGAGGAGTACATGGTACAGCTTGAAGAACATTT 480  
QY 1170 TCAGCTTCTCGCAGCAGAGATGCAAGTGGATAAAAAGAGATATATCTGGCTACTGATGA 1229  
Db 481 TCAGCTTCTCGCAGCAGAGAT-----501  
QY 1230 TCCTACTTTGTTAAAGGAGGCAAGACAAAGTACTCCAATTATGAAATTTATTTAGTGATAA 1289  
Db 502 -----GTACTCCAATTATGAAATTTATTTAGTGATAA 531  
QY 1290 CTCATTTCTTGGTCAGCTGAGTACACAAATCGGTACAGAGAAATTCACCTCGGGGTG 1349  
Db 532 CTCATTTCTTGGTCAGCTGAGTACACAAATCGGTACAGAGAAATTCACCTCGGGGTG 591  
QY 1350 GATCCTGGATATACATTTCTCTCACAGGCTGACTTTCTAGTGTGACTTTTTCATCCCA 1409  
Db 592 GATCCTGGATATACATTTCTCTCACAGGCTGACTTTCTAGTGTGACTTTTTCATCCCA 651  
QY 1410 GGTCTGTGGTGTGCTTATGAAATCATGCAAAACCTGCAATCCTGATGCTCTGCGAACTT 1469  
Db 652 GGTCTGTGGTGTGCTTATGAAATCATGCAAAACCTGCAATCCTGATGCTCTGCGAACTT 711  
QY 1470 CAACTTTTGGATGATCACTACTATTTGGAGGCCAAATGSCCAATCAGATTCGCTGT 1529  
Db 712 -CAITCTTTGGATGATCACTACTATTTGGAGGCCAAATGSCCAATCAGATTCGCTGT 770  
QY 1530 TTATCTCTCAAAACC-TCGAACTGAAGA-GGAAATTTCCAATGGAACCTGGAGATAT 1583  
Db 771 TTATCTCTCAAAACCTTCGAACTGAAGAGGGAATTCNAATGGAACCTGGAGAT 826

## RESULT 12

BX450010

LOCUS

DEFINITION BX450010 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA clone CS0DG005Y007 5-PRIME, mRNA sequence.

ACCESSION

BX450010

VERSION

BX450010.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7499.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0AS0092B08QPl&amp;cluster=7499.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0AS0092B08QPl.

Location/Qualifiers

1. .1027

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DG005Y007"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/clone\_line="RAMOS CELL LINE"

/note="Vector: pcMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pcMVSPORT 6 vector. Library was not normalized."

## FEATURES

source

BASE COUNT 320 a 188 c 258 g 252 t 9 others  
ORIGIN  
Query Match 40.3%; Score 696.8; DB 13; Length 1027;  
Best Local Similarity 92.4%; Pred. No. 8.6e-170;  
Matches 728; Conservative 4; Mismatches 56; Indels 0; Gaps 0;  
QY 1 ATGCGGCATGAGTGGTTCCTGGCGTTGGATATGCTCATCTTTTTCCTCGGGGACC 60  
Db 240 ATGCGGCATGAGTGGTTCCTGGCGTTGGATATGCTCATCTTTTTCCTCGGGGACC 299  
QY 61 TTGTATTATATAGTGGTTCATTTGTTTCGAGATATGACACCCCTGATCACTCCAGC 120  
Db 300 TTGTGTTTTATATAGTGGTTCATTTGTTTCGAGATATGACACCCCTGATCACTCCAGC 359  
QY 121 AGAGAACTCTCCAAAGATTTCTTGAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 180  
Db 360 CGAGAACTGTCCAAGATTTCTGCAAGCTTGAACGCTTAAACAGCAAAATGAAGACTTG 419  
QY 181 AGGCGAATGGTGTAGTCTCTCCGAATACCGAAGGCCCCATTGACCGGGGACAGCTACA 240  
Db 420 AGGCGAATGGGCGAATCTCTCCGATACCGAAGGCCCCATTGATGATCGGGGCCAGCTATA 479  
QY 241 GGAAGAGTCCGTGTTTTAGAGAGACAGCTTGTAAAGCCAAAGACAGATTTGAAAATTTAC 300  
Db 480 GGAAGAGTACCGCTTTTAGAAGAGCAGCTTGTAAAGCCAAAGACAGATTTGAAAATTTAC 539  
QY 301 AAGAAAACAAGCTAGAAATGTCTGGGAAGGATCATGAAATCTTTAAGAGAGGAGTTGAA 360  
Db 540 AAGAAAACAGACAGAAATGTCTGGGAAGGATCATGAAATCTCTGAGGAGGAGTTGAA 599  
QY 361 AATGAGCTAAGAGCTCTGTTTTTCTCAAGCGAAGCTTGAAGAAATTTAAGCAATTTA 420  
Db 600 AATGAGCTAAGAGCTCTGTTTTTCTCAAGCGAAGCTTGAAGAAATTTAAGCAATTTA 659  
QY 421 GAAAGAAATGAACTCCAAAGACATCGAGATGAAATCTTTTGGATTTAGGACACCATGAA 480  
Db 660 GAAAGAAATGAACTCCAAAGACATCGAGATGAAATCTTTTGGATTTAGGACATCATGAA 719  
QY 481 AGGTCTATCATGACAGATCTATCTACTCTCAGTCAACAGATGGAGCAGGGGATTTGGCGT 540  
Db 720 AGGTCTATATGACGGATCTATCTACTCTCAGTCAACAGATGGAGCAGGGGATTTGGCGG 779  
QY 541 GAAAGAGGCGCAAGATCTGACAGAGCTGTGTCAGCGGAGAAATTAACATATCTCCAGAA 600  
Db 780 GAAAGAGGCGCAAGATCTGACAGAACTGTTTCAGCGGAGAAATTAACATATCTCCAGAA 839  
QY 601 CCTAAGGACTGACGAGCAAGCCAGGAGCTGTGTGAACATCAATAAAGGCTGTGGCTAT 660  
Db 840 CCCAAGGACTGACGAGCAAGCCAAAGCTGTGTGAATATCAACAAAGGCTGTGGCTAT 899  
QY 661 GGTGTCAACTCCATCAGTGTCTACTGTTTTCATGATTTGCTTATGGCACCAGCGAACA 720  
Db 900 GGTGTCAAGTCCATCATGTGCTCTACTGTTTCATGATTTGCTTATGGCACCAGCGAACA 959  
QY 721 CTCATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGCTGTGATGGAGACATGTTTGA 780  
Db 960 CTCATCTTGAATCTCAGAAATTTGGCGCTATGCTACTGCGKGGATGGAGATGTTATTTAGG 1019  
QY 781 CCTGTAG 788  
Db 1020 TCTGTAG 1027

## RESULT 13

AUI24128

LOCUS

DEFINITION

AUI24128 NT2RM2 Homo sapiens cDNA clone NT2RM2001714 5', mRNA sequence.

ACCESSION

AUI24128

VERSION

AUI24128.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

AUI24128 834 bp mRNA linear EST 01-AUG-2002  
AUI24128 NT2RM2 Homo sapiens cDNA clone NT2RM2001714 5', mRNA sequence.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 834)  
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
Y., Sugano,S., Isogai,T.)  
JOURNAL Unpublished  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
FEATURES  
Location/Qualifiers  
1..834  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2PM2001714"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/clone\_lib="NT2RM2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"  
BASE COUNT 249 a 187 c 181 g 215 t 2 others  
ORIGIN  
Query Match 39.8%; Score 687.4; DB 9; Length 834;  
Best Local Similarity 90.4%; Pred. No. 2.2e-167;  
Matches 744; Conservative 0; Mismatches 78; Indels 1; Gaps 1;  
QY 905 CTATTACCTGGCTGTTCCAGAAAGACCTTGCAGACCGACTCTAGAGTCCATGGTG 964  
DB 1 CATATTATACCTTGGCTGTACCAAGAGACCTCGCAGATCGACTTGTACGAGTGCATGGTG 60  
QY 965 ACCCTGCAGTGTGGGTGTCCTCCAGTTGTCAATATCTGATTCCTCCACAACTTCGCG 1024  
DB 61 ACCCTGCAGTGTGGGTGTCCTCCAGTTGTCAATATCTGATTCCTCCACAACTTCGCG 120  
QY 1025 TGGAAAGGAATAGAAAGACCCACCAAGAGCTTGGCTTCAAAATCCAGTTATTGGAG 1084  
DB 121 TAGAAAGGAATAGAAAGACCCACCAAGAGCTTGGCTTCAAAATCCAGTTATTGGAG 180  
QY 1085 TCCATGTCCAGCGCAGACAAAGTGGGAACAGAGCGCTTCCACCCCATCGAGGAGT 1144  
DB 181 TCCATGTCCAGCGCAGACAAAGTGGGAACAGAGCGCTTCCATCCCATTTGAAGAGT 240  
QY 1145 ACATGGTACAGTTGAAGAACATTTTCAGCTTCTCGACGAGATGCAAGTGGATATAA 1204  
DB 241 ACATGGTACAGTTGAAGAACATTTTCAGCTTCTCGACGAGATGCAAGTGGATATAA 300  
QY 1205 AAAGAGTATATCTGCTACTGATGATCTTCTTAAAGAGGCAAGACAAAGTACT 1264  
DB 301 AAAGAGTATATCTGCTACTGATGATCTTCTTAAAGAGGCAAGACAAAGTACT 360  
QY 1265 CCAATTATGAATTTATTAGTGATACTCTATTCTTGGTCAGCTGGACTACCAATCGGT 1324  
DB 361 CCAATTATGAATTTATTAGTGATACTCTATTCTTGGTCAGCTGGACTGGACCAATCGAT 420  
QY 1325 ACACAGAAAATTCACCTCGGGGTGTGATCTCGGATATACATTTCTCTCACAGGCTGACT 1384  
DB 421 ACACAGAAAATTCACCTCGGGGTGTGATCTCGGATATACATTTCTCTCACAGGCTGACT 480  
QY 1385 TTCTAGTGTGACTTTTTCATCCCGAGTCTGTCGGGTGCTTATGAATCATGCAAAACC 1444

Db 481 TCCTAGTGTGACTTTTTCATCCCGAGTCTGTGAGTTGCTATGAAATATGCAAAAC 540  
QY 1445 TGCATCTGATGCTCTGCGAACTTCCATTCTTTGGATGACATCTACTATTTCGAGGCC 1504  
Db 541 TACGTCTCTGATGCTCTGCGAACTTCCATTCTTTAGATGACATCTACTATTTCGAGGCC 600  
QY 1505 AAAATGCCCAATCAAGTATGCTGTTTATCTCTCAAAACCTCGAACTGGAAGAGAAATTC 1564  
Db 601 AGAATGCCCAATCAATCAATTTGCAATTTATGCTCAACCAACCCGAACTGCAGATGAAATTC 660  
QY 1565 CAATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTTATTCTAAAG 1624  
Db 661 CCATGGAACCTGGAGATATCATTTGGTGTGCTGGAACCATTTGGATGGTTATTCTAAAG 720  
QY 1625 GTATCAACAGAAAACCTTGGAAAACAGCTTATATCTCTCTCAAAAGTCCGAGAGAAGA 1684  
Db 721 GTGTCAACAGAAAATTTGGGAAGACGCGCTATAT-CTCTCAAAAGTTCGAGAGAAGA 779  
QY 1685 TAGAAACAGTCAAGTATCCACATATCTGAAGCTGAAAATA 1727  
Db 780 TAGAAACGTCAGTATCCCAATATCTCTGAGCTGAGAAAATA 822  
RESULT 14  
BG973221 702 bp mRNA linear EST 12-JUN-2001  
LOCUS 602842563f1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:4978383 5',  
DEFINITION mRNA sequence.  
ACCESSION BG973221  
VERSION BG973221.1 GI:14360858  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10974 row: n column: 16  
High quality sequence stop: 694.  
FEATURES  
Location/Qualifiers  
1..702  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NMRI"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4978383"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam4"  
/note="Oxgan: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Priscilla Furth,  
NIH Reference for transgenic model: Li et al., Cell Growth  
and Differentiation 7, 3-11 (1996)."  
BASE COUNT 217 a 145 c 174 g 166 t  
ORIGIN  
Query Match 38.3%; Score 660.2; DB 12; Length 702;









**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 38.5 Seconds  
(without alignments)

2370.592 Million cell updates/sec

Title: US-09-971-773-23

Perfect score: 3081

Sequence: 1 MRAWTGSWRWIMLILFAWGT.....YKVRKTIETVKYPTYPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3081	100.0	575	23 ABG34135	Antibody productio
2	3024	98.1	575	23 ABG34136	Antibody productio
3	2984	96.9	575	18 AAW22125	Human alpha 1-6 fu
4	2884	96.9	575	23 ABB08405	Alpha1,6-fucosyl t
5	2885	93.6	575	18 AAW22124	Pig alpha 1-6 fuco
6	2639	85.7	515	22 AAG73884	Human colon cancer
7	1771	57.5	339	22 AAB75061	Human alpha 1-6 fu
8	1489.5	48.3	619	22 ABB59117	Drosophila melanog
9	1193	38.7	233	22 AAB75062	Human alpha 1-6 fu

10	431	14.0	82	22	ABG48804	Human liver peptid
11	431	14.0	82	22	ABB28797	Peptide #1448 enco
12	431	14.0	82	22	ABB33985	Peptide #1491 enco
13	431	14.0	82	22	ABB19422	Protein #1421 enco
14	431	14.0	82	22	AAW54747	Human brain expres
15	431	14.0	82	22	AAW67139	Human bone marrow
16	431	14.0	82	22	AAW15003	Peptide #1437 enco
17	431	14.0	82	22	AAW27440	Peptide #1477 enco
18	431	14.0	82	22	AAW02730	Peptide #1412 enco
19	431	14.0	82	23	ABG36801	Human peptide enco
20	425	13.8	81	22	AAW64859	Human brain expres
21	425	13.8	81	22	AAW37789	Peptide #11826 enc
22	133	4.3	1959	23	ABJ10604	Human novel protei
23	131	4.3	25	18	AAW22131	Fragment #2 of hum
24	130.5	4.2	751	22	ABW62477	Drosophila melanog
25	124.5	4.0	1960	22	AAW78854	Human protein SEQ
26	124.5	4.0	2143	22	ABG01716	Novel human diagno
27	124	4.0	1025	23	ABB90967	Herbicide activ
28	124	4.0	1285	24	ABR47540	Breast cancer asso
29	124	4.0	2442	21	AAW77575	Human cytoskeletal
30	123	4.0	1963	22	AAW79838	Human protein SEQ
31	121.5	3.9	898	24	AAE33671	Human structural a
32	121.5	3.9	1014	23	ABG61923	Prostate cancer-as
33	120.5	3.9	3542	22	AAW62142	P. faicparum FCR3
34	119.5	3.9	612	22	AAW95546	Human protein sequ
35	118.5	3.8	976	22	AAW66581	Human SCP-1 mutain
36	118.5	3.8	976	24	ABP74709	Human SCP-1 protei
37	117.5	3.8	1427	12	AAW10534	Human 160kD mediat
38	116.5	3.8	931	22	AAW79504	Human protein SEQ
39	116.5	3.8	990	22	AAW78520	Human protein SEQ
40	116	3.8	214	18	AAW09037	Osteoclast stimula
41	116	3.8	214	18	AAW12706	Osteoclast stimula
42	116	3.8	229	17	AAW05403	Human clone 5 prot
43	116	3.8	283	21	AAW28744	Arabidopsis thalia
44	116	3.8	2383	23	ABG56531	Human breast speci
45	115.5	3.7	888	23	AAW83013	Human homologue of

#### ALIGNMENTS

#### RESULT 1

ABG34135  
ID ABG34135 standard; Protein; 575 AA.

AC ABG34135;

XX 15-JUL-2002 (first entry)

DT Antibody production method related protein #1.

DE Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW antibacterial; antinflammatory; antiallergic; allergy; inflammation;  
KW autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW tumour; circulatory disease; infection.

XX Cricetulus griseus.

OS WO200231140-A1.

PN 18-APR-2002.

PD 05-OCT-2001; 2001WO-JP08804.

XX 06-OCT-2000; 2000JP-0308526.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

XX Kanda Y, Sato M, Nakamura K, Uchida K, Shinkawa T, Yamane N;

XX Hosaka E, Yamano K, Yamasaki M, Hanai N;

XX WPI; 2002-340182/37.

```
PT Cells producing antibody compositions including antibody fragments and
PT fusion proteins with Fc domain of antibody, useful for prevention or
PT treatment of cancer, immune diseases, circulatory diseases and
PT infections
XX
XX Claim 23; Page 18-21; 314pp; Japanese.
XX
XX This invention relates to novel method for antibody production
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell
XX transferred with a gene encoding an antibody molecule for producing a
XX composition comprising an antibody molecule with an Fc domain bonded
XX to the N-glycoside linkage complex sugar chain. The produced antibody
XX compositions are drugs for prevention or treatment of diseases
XX accompanying tumour, allergy or inflammation, autoimmune diseases,
XX circulatory diseases, and viral and bacterial infections. The
XX antibodies can be stably produced using the method of the invention
XX with high binding activity and potency thus leading to high safety and
XX reduced side effects when applied alone or in combination with other
XX drugs for therapy. The present sequence represents a protein
XX molecule used in the method of the invention.
XX
XX Sequence 575 AA;
XX
XX Query Match 100.0%; Score 3081; DB 23; Length 575;
XX Best Local Similarity 100.0%; Pred. No. 5.5e-272;
XX Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRELSKILAKLERLKQONEDL 60
DB 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRELSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEIQENYKQARNDLGKHILRRRIE 120
DB 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEIQENYKQARNDLGKHILRRRIE 120
QY 121 NGAKELWFFLQSELKKLKEGNEQLVKAKEIQENYKQARNDLGKHILRRRIE 180
DB 121 NGAKELWFFLQSELKKLKEGNEQLVKAKEIQENYKQARNDLGKHILRRRIE 180
QY 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYCOLHHVVCFMIAVGTORT 240
DB 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYCOLHHVVCFMIAVGTORT 240
QY 241 LILESQNRWYATGWTETPRVPSVETCTDRSGLSGHWSEVKDKNVQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGWTETPRVPSVETCTDRSGLSGHWSEVKDKNVQVVELPIVDSLHPR 300
QY 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHVI 360
DB 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHVI 360
QY 361 GVHVRRTDKVGEAAFPHPLEEVYVVEHFQLLERMMKVKRVLATDDPSLLKEATK 420
DB 361 GVHVRRTDKVGEAAFPHPLEEVYVVEHFQLLERMMKVKRVLATDDPSLLKEATK 420
QY 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYETMQ 480
DB 421 YSNYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYETMQ 480
QY 481 TLHPDASANFHLDDIYFPGGQNAHQIAVYPHQRTKEEIPMEPGDIIIGVAGNHNWNGYS 540
DB 481 TLHPDASANFHLDDIYFPGGQNAHQIAVYPHQRTKEEIPMEPGDIIIGVAGNHNWNGYS 540
QY 541 KGVNRKLGKTLGVPSYKREKLETVKYPTYPEAK 575
DB 541 KGVNRKLGKTLGVPSYKREKLETVKYPTYPEAK 575
XX
XX RESULT 2
XX ABG34136
XX ID ABG34136 standard; Protein; 575 AA.
XX
XX AC ABG34136;
```

```
XX
XX 15-JUL-2002 (first entry)
XX
XX Antibody production method related protein #1.
XX
XX Antibody production; cytostatic; immunomodulator; vasotropic; virucide;
XX antibacterial; antiinflammatory; antiallergic; allergy; inflammation;
XX autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;
XX tumour; circulatory disease; infection.
XX
XX Mus musculus.
XX
XX WO200231140-A1.
XX
XX 18-APR-2002.
XX
XX 05-OCT-2001; 2001WO-JP08804.
XX
XX 06-OCT-2000; 2000JP-0308526.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
XX WPI; 2002-340182/37.
XX
XX Cells producing antibody compositions including antibody fragments and
XX fusion proteins with Fc domain of antibody, useful for prevention or
XX treatment of cancer, immune diseases, circulatory diseases and
XX infections
XX
XX Claim 33; Page 21-23; 314pp; Japanese.
XX
XX This invention relates to novel method for antibody production
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell
XX transferred with a gene encoding an antibody molecule for producing a
XX composition comprising an antibody molecule with an Fc domain bonded
XX to the N-glycoside linkage complex sugar chain. The produced antibody
XX compositions are drugs for prevention or treatment of diseases
XX accompanying tumour, allergy or inflammation, autoimmune diseases,
XX circulatory diseases, and viral and bacterial infections. The
XX antibodies can be stably produced using the method of the invention
XX with high binding activity and potency thus leading to high safety and
XX reduced side effects when applied alone or in combination with other
XX drugs for therapy. The present sequence represents a protein
XX molecule used in the method of the invention.
XX
XX Sequence 575 AA;
XX
XX Query Match 98.1%; Score 3024; DB 23; Length 575;
XX Best Local Similarity 97.6%; Pred. No. 8.8e-267;
XX Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;
XX
QY 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRELSKILAKLERLKQONEDL 60
DB 1 MRAWTGSRWIMLILFAWGTLIFYGGHLVRNDHPDHSRELSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEIQENYKQARNDLGKHILRRRIE 120
DB 61 RMAESLRIPGPIDOGTATGRVRLVEQLVKAKEIQENYKQARNDLGKHILRRRIE 120
QY 121 NGAKELWFFLQSELKKLKEGNEQLVKAKEIQENYKQARNDLGKHILRRRIE 180
DB 121 NGAKELWFFLQSELKKLKEGNEQLVKAKEIQENYKQARNDLGKHILRRRIE 180
QY 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYCOLHHVVCFMIAVGTORT 240
DB 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYCOLHHVVCFMIAVGTORT 240
QY 241 LILESQNRWYATGWTETPRVPSVETCTDRSGLSGHWSEVKDKNVQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGWTETPRVPSVETCTDRSGLSGHWSEVKDKNVQVVELPIVDSLHPR 300
```

QY 301 PPYLPLAVPEDLADLLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360  
 DB 301 PPYLPLAVPEDLADLLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360  
 QY 361 GVHVRTDKVGTAAAFHPPIEYVMVHVEEHFOLLERRMKVDKRVYLAATDDPSLKEAKTK 420  
 DB 361 GVHVRTDKVGTAAAFHPPIEYVMVHVEEHFOLLARRMQDKRVYLAATDDPSLKEAKTK 420  
 QY 421 YSNYEFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480  
 DB 421 YSNYEFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNWGS 540  
 DB 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNWGS 540  
 QY 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575  
 DB 541 KGINRKLKGTGLYPSYKVKREKIEYKTYPTYPEAEK 575

## RESULT 3

AAW22125  
 ID AAW22125 standard; Protein; 575 AA.

AC AAW22125;

DT 05-MAR-1998 (first entry)

DE Human alpha 1-6 fucosyltransferase.

KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 KW Guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GlnAC; cancer diagnosis.

OS Homo sapiens.

PN WO9727303-A1.

PD 31-JUL-1997.

PF 23-JAN-1997; 97WO-JP00171.

PR 22-JUL-1996; 96JP-0192260.

PR 24-JAN-1996; 96JP-0010365.

PR 21-JUN-1996; 96JP-0161648.

PR 24-JUN-1996; 96JP-0162813.

PA (TOYM ) TOYO BOSEKI KK.

PI Shiba T, Taniguchi N, Uozumi N, Yanagidani S;

DR WPI; 1997-393690/36.

DR N-PSDB; AAT76574.

XX Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 PT synthesis and modification of sugar chains and used as an antigen  
 PT for production of diagnostic antibodies

PS Claim 17; Page 39-43; 61pp; Japanese.

XX AAW22124 and AAW22125 represent the pig and human alpha 1-6  
 CC fucosyltransferases of the invention, respectively. The enzyme transfers  
 CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlnAC  
 CC nearest to R in the receptor molecule: (GlnACbeta 1-2Manalpha 1-6)  
 CC (GlnACbeta 1-2Manalpha 1-3)Manbeta 1-4GlnACbeta 1-4GlnAC-R to give  
 CC (GlnACbeta 1-2Manalpha 1-6)(GlnACbeta 1-2Manalpha 1-3)Manbeta  
 CC 1-4GlnACbeta 1-4(Fucalpha 1-6)(GlnAC-R. It has an optimum pH of about  
 CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
 CC hours at 4 degrees C. The optimum working temperature of the  
 CC alpha 1-6 fucosyltransferases is 30-37 degrees C. A bivalent metal is  
 CC not required for activity of the enzyme, and the enzyme is not inhibited

CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 CC modification of sugar chains, and as antigen for the production of  
 CC antibodies recognising the enzyme. The antibodies can be used for the  
 CC diagnosis of cancer and other diseases.

SQ Sequence 575 AA;

Query Match 96.9%; Score 2984; DB 18; Length 575;

Best Local Similarity 96.2%; Pred. No. 3.9e-263;

Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRATGSGWRMTMLILFWAGTLLFYIGGHLVRDNDHPDHSRELSKILAKLRLKQONEDL 60

DB 1 MRPTGSGWRMTMLILFWAGTLLFYIGGHLVRDNDHPDHSRELSKILAKLRLKQONEDL 60

QY 61 RMAESLRIPGPDQGTATGRVRVLEEQVKAEQIENYKQARNDLGKDEILRRRIE 120

DB 61 RMAESLRIPGPDQGTATGRVRVLEEQVKAEQIENYKQARNDLGKDEILRRRIE 120

QY 121 NGAKELWFFLQSELKLLKLEGNELQRHADILLDLGHHERSIMTDLVYLSQTDGAGWR 180

DB 121 NGAKELWFFLQSELKLLKLEGNELQRHADILLDLGHHERSIMTDLVYLSQTDGAGWR 180

QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNINKGCGYCOLHHVVCYCMIAYGTOPT 240

DB 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNINKGCGYCOLHHVVCYCMIAYGTOPT 240

QY 241 LILESQNRVATGWTVFVRVSETCTDRSGLSTGHSGEVKDNQVQVVELPIVDSLHPR 300

DB 241 LILESQNRVATGWTVFVRVSETCTDRSGLSTGHSGEVKDNQVQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADLLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360

DB 301 PPYLPLAVPEDLADLLRVHGDPAVWVVSQFVKYLIRPQWLERIEETTKLGFKEHPVI 360

QY 361 GVHVRTDKVGTAAAFHPPIEYVMVHVEEHFOLLERRMKVDKRVYLAATDDPSLKEAKTK 420

DB 361 GVHVRTDKVGTAAAFHPPIEYVMVHVEEHFOLLARRMQDKRVYLAATDDPSLKEAKTK 420

QY 421 YSNYEFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480

DB 421 YSNYEFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADPLVCTFSSQVCRVAYEIMQ 480

QY 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNWGS 540

DB 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHOPRTKEEIPMEPGDIIGVAGNHNWGS 540

QY 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575

DB 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575

## RESULT 4

ABB08405

ID ABB08405 standard; Protein; 575 AA.

XX ABB08405;

DT 07-MAY-2002 (first entry)

DE Alpha1,6-fucosyl transferase amino acid sequence.

KW Plant; glycoprotein; alpha1,6-fucosyl transferase; alpha1,6-Ft;  
 KW enzyme.

OS Homo sapiens.

PN JP2001333787-A.

PD 04-DEC-2001.

PF 06-MAR-2001; 2001JP-0063704.

XX

PR 22-MAR-2000; 2000JP-0081059.  
 XX (TANI/) TANIGUCHI N.  
 PA (SEKI/) SEKI T.  
 XX (FUJI/) FUJIYAMA K.  
 XX  
 DR WPI; 2002-158816/21.  
 DR N-PSDB; ABA98809.  
 XX  
 XX A plant cell with an animal type sugar chain adding function, for the  
 PT preparation of a glycoprotein with an animal type sugar chain -  
 XX  
 PS Example 1; Page 24-26; 38pp; Japanese.  
 XX  
 CC The invention relates to a plant cell with an animal type sugar chain  
 CC adding function, created by transforming a tobacco cell with a gene  
 CC encoding an enzyme derived from an animal which can transfer a fucose  
 CC residue to the reductive end acetylglucosamine residue of a sugar chain.  
 CC The gene that is introduced into the plant cell encodes the enzyme  
 CC alpha1,6-fucosyl transferase. The method of the invention is useful for  
 CC the preparation of a glycoprotein having animal type sugar chain. The  
 CC current sequence represents alpha1,6-fucosyl transferase amino acid  
 CC sequence.  
 XX  
 XX Sequence 575 AA;  
 SQ  
 Query Match 96.9%; Score 2984; DB 23; Length 575;  
 Best Local Similarity 96.2%; Pred. No. 3.9e-263;  
 Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRDNDHPDSSRELKILAKLERLKKQNEDL 60  
 Db 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRDNDHPDSSRELKILAKLERLKKQNEDL 60  
 QY 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDL-GKDHILRRRIE 120  
 Db 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDL-GKDHILRRRIE 120  
 QY 121 NGAKELWFFLQSELKKLKNLEGNLQKHADEILDLGHHERSIMTDLYYLSQTDGAGWR 180  
 Db 121 NGAKELWFFLQSELKKLKNLEGNLQKHADEILDLGHHERSIMTDLYYLSQTDGAGWR 180  
 QY 181 EKEAKDLTBLVORRTYLLQNPDKCSKARKLVNINKGCGYQGLHHVYVCFMIYGTORT 240  
 Db 181 EKEAKDLTBLVORRTYLLQNPDKCSKARKLVNINKGCGYQGLHHVYVCFMIYGTORT 240  
 QY 241 LILESQNWRYATGGWETVFRPYSETCTDRSGLSTGHWGSEVKDKNVQVVELPIVDSLHPR 300  
 Db 241 LILESQNWRYATGGWETVFRPYSETCTDRSGLSTGHWGSEVKDKNVQVVELPIVDSLHPR 300  
 QY 301 PPYLPLAVPEDLADRLRLRHGDPVWVWSQFVKYLIRPQWLERIEIETTKLGFKHPVI 360  
 Db 301 PPYLPLAVPEDLADRLRLRHGDPVWVWSQFVKYLIRPQWLERIEIETTKLGFKHPVI 360  
 QY 361 GVHVRTDKVGTAAFPPIETVMVHVEHFOLLERMMKVKRKYLATDDPSLLKEATK 420  
 Db 361 GVHVRTDKVGTAAFPPIETVMVHVEHFOLLERMMKVKRKYLATDDPSLLKEATK 420  
 QY 421 YSNYEFISDNTSWAGLHNRVTENSLRGVILDIHFLSQADPLVCTFSQCVRAVEIMQ 480  
 Db 421 YSNYEFISDNTSWAGLHNRVTENSLRGVILDIHFLSQADPLVCTFSQCVRAVEIMQ 480  
 QY 481 TLHPDASANFHSLLDIYFGGQNAJNIAVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540  
 Db 481 TLHPDASANFHSLLDIYFGGQNAJNIAVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540  
 QY 541 KGVNRKLGTLGTPSYKVKREKTIETVKYPTYPEAK 575  
 Db 541 KGVNRKLGTLGTPSYKVKREKTIETVKYPTYPEAK 575

RESULT 5  
 AAW22124

ID AAW22124 standard; Protein; 575 AA.  
 XX  
 AC AAW22124;  
 XX  
 DT 05-MAR-1998 (first entry)  
 XX  
 DE Pig alpha 1-6 fucosyltransferase.  
 XX  
 KW Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;  
 KW guanosine diphosphate; sugar chain synthesis; modification; antibody;  
 KW GlcNAc; cancer diagnosis.  
 XX  
 OS Sus scrofa.  
 XX  
 PN W09727303-A1.  
 XX  
 PD 31-JUL-1997.  
 XX  
 PF 23-JAN-1997; 97WO-JP00171.  
 XX  
 PR 22-JUL-1996; 96JP-0192260.  
 PR 24-JAN-1996; 96JP-0010365.  
 PR 21-JUN-1996; 96JP-0161648.  
 PR 24-JUN-1996; 96JP-0162813.  
 XX  
 XX (TOYM) TOYO BOSEKI KK.  
 PA  
 PI Shiba T, Taniguchi N, Uozumi N, Yanagidani S;  
 XX  
 DR WPI; 1997-393690/36.  
 DR N-PSDB; AAT76573.  
 XX  
 PT Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for  
 PT synthesis and modification of sugar chains and used as an antigen  
 PT for production of diagnostic antibodies  
 XX  
 PS Claim 4; Page 30-34; 61pp; Japanese.  
 XX  
 CC AAW22124 and AAW22125 represent the pig and human alpha 1-6  
 CC fucosyltransferases of the invention, respectively. The enzyme transfers  
 CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc  
 CC nearest to R in the receptor molecule: (GlcNAc-beta 1-2-Man-alpha 1-6)  
 CC (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta 1-4-GlcNAc-beta 1-4-GlcNAc-R to give  
 CC (GlcNAc-beta 1-2-Man-alpha 1-6) (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta  
 CC 1-4-GlcNAc-beta 1-4 (Fucal-alpha 1-6) GlcNAc-R. It has an optimum pH of about  
 CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5  
 CC hours at 4 degrees C. The optimum working temperature of the  
 CC alpha 1-6 fucosyltransferase is 30-37 degrees C. A bivalent metal is  
 CC not required for activity of the enzyme, and the enzyme is not inhibited  
 CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and  
 CC modification of sugar chains, and as antigen for the production of  
 CC antibodies recognising the enzyme. The antibodies can be used for the  
 CC diagnosis of cancer and other diseases.  
 XX  
 SQ Sequence 575 AA;  
 Query Match 93.6%; Score 2885; DB 18; Length 575;  
 Best Local Similarity 92.9%; Pred. No. 4.3e-254;  
 Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRDNDHPDSSRELKILAKLERLKKQNEDL 60  
 Db 1 MRWGTGSRWIMLILFWAGTLLFYGGHLVRDNDHPDSSRELKILAKLERLKKQNEDL 60  
 QY 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDL-GKDHILRRRIE 120  
 Db 61 RMAESLRIPGPIDOGTATGRVRLVLEOLVKAKEQIENYKKAQANDL-GKDHILRRRIE 120  
 QY 121 NGAKELWFFLQSELKKLKNLEGNLQKHADEILDLGHHERSIMTDLYYLSQTDGAGWR 180  
 Db 121 NGAKELWFFLQSELKKLKNLEGNLQKHADEILDLGHHERSIMTDLYYLSQTDGAGWR 180  
 QY 181 EKEAKDLTBLVORRTYLLQNPDKCSKARKLVNINKGCGYQGLHHVYVCFMIYGTORT 240

Db 181 EKEARDLTQLVQRRTYLLQNPDKCSKAKKLVNINKGCGYGCQLHHVYVCFMAYGTQRT 240  
Qy 241 LILESQWRYATGQWETVFRPVSECTDRSGLSGHSWGEVKNQVQVVELPIVDSLHPR 300  
Db 241 LALESHNNRYATGQWETVFRPVSECTDRSGLSGHSWGEVKNQVQVVELPIVDSVHPR 300  
Qy 301 PPYLPLAVPEDLADRLVHGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGKHPVI 360  
Db 301 PPYLPLAVPEDLADRLVHGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGKHPVI 360  
Qy 361 GVHVRTDKVTEAAPHPIEYMWVHVEHFOLERRMKVDKRVYLAATDDPSLLKEATK 420  
Db 361 GVHVRTDKVTEAAPHPIEYMWVHVEHFOLERRMKVDKRVYLAATDDPSLLKEATK 420  
Qy 421 YSNVEFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYELMQ 480  
Db 421 YPSVEFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYELMQ 480  
Qy 481 TLHPDASANFSLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWGYS 540  
Db 481 ALHPDASANFSLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWGYP 540  
Qy 541 KGVNRKLGKTGLPSYKVKREKIEVYKTYPTPEAEK 575  
Db 541 KGVNRKLGKTGLPSYKVKREKIEVYKTYPTPEADK 575

## RESULT 6

AAG73884  
ID AAG73884 standard; Protein; 515 AA.

XX AC AAG73884;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:4648.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
colorectal carcinoma; chromosome 14.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357/24.

XX DR N-P5DB; AAH33315.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
useful for preventing, diagnosing and/or treating colorectal cancers -

XX PS Claim 11; Page 6451-6453; 9803pp; English.

XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
cancer-associated nucleic acid molecules (N) and proteins (P), where  
the proteins are collectively known as colon cancer antigens. The colon  
cancer antigens have cytostatic activity and can be used in gene  
therapy and vaccine production. N and P may be used in the prevention,  
diagnosis and treatment of diseases associated with inappropriate P  
expression. For example, N and P may be used to treat disorders  
associated with decreased expression by rectifying mutations or deletions  
in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 515 AA;

Query Match 85.7%; Score 2639; DB 22; Length 515;

Best Local Similarity 95.7%; Pred. No. 1e-231;

Matches 489; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 65 ESLRIPGPDIDQGTATGRVRLERQVLKAKEIQENYKQARNDLGKHIEILRRRIENGAK 124

Db 5 KSLRIPGPDIDQGTATGRVRLERQVLKAKEIQENYKQARNDLGKHIEILRRRIENGAK 64

Qy 125 ELWFFLOSELKLLKLGNEIQRHADELILDLGHHSIMTDLYLVSQDAGDWREKEA 184

Db 65 ELWFFLOSELKLLKLGNEIQRHADELILDLGHHSIMTDLYLVSQDAGDWREKEA 124

Qy 185 KDLTELQRRITTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVYVCFMAYGTQRTLILE 244

Db 125 KDLTELQRRITTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVYVCFMAYGTQRTLILE 184

Qy 245 SQNWRYATGQWETVFRPVSECTDRSGLSGHSWGEVKNQVQVVELPIVDSLHPRPPYL 304

Db 185 SQNWRYATGQWETVFRPVSECTDRSGLSGHSWGEVKNQVQVVELPIVDSLHPRPPYL 244

Qy 305 PLAYPEDLADRLVHGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGKHPVIGVHV 364

Db 245 PLAYPEDLADRLVHGDPAVWVVSQFVKYLIRPQWLERIEETTKKLGKHPVIGVHV 304

Qy 365 RRTDKVGTAAAFHPIEYMWVHVEHFOLERRMKVDKRVYLAATDDPSLLKEATKYSNY 424

Db 305 RRTDKVGTAAAFHPIEYMWVHVEHFOLERRMKVDKRVYLAATDDPSLLKEATKYPNY 364

Qy 425 EFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYELMQTLHP 484

Db 365 EFISDNTSISWAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYELMQTLHP 424

Qy 485 DASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWGYSKGVN 544

Db 425 DASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHWGYSKGVN 484

Qy 545 RKLKGTGLPSYKVKREKIEVYKTYPTPEAEK 575

Db 485 RKLKGTGLPSYKVKREKIEVYKTYPTPEAEK 515

## RESULT 7

AAH75061

ID AAB75061 standard; Protein; 339 AA.

XX AC AAB75061;

XX DT 20-JUL-2001 (first entry)

XX DE Human alpha 1-6 fucosyltransferase protein 237-575 SEQ ID NO:3.

XX KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;

XX OS Homo sapiens.

XX PN JP2001011097-A.

XX PD 16-JAN-2001.

XX XX







```
ABB28797
ID ABB28797 standard; Peptide; 82 AA.
XX
AC ABB28797;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #1448 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
XX
KW disease; cancer.
XX
XX Homo sapiens.
XX
OS WO200157271-A2.
XX
PN 09-AUG-2001.
XX
PD
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
DR
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX
PS Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 82 AA;
Query Match 14.0%; Score 431; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-31;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 280 EVKDKNQVVELPIVDSLHPRPPYLPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
DB 1 EVKDKNQVVELPIVDSLHPRPPYLPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60

QY 340 PWLEREIEETTKLGFKHPVIG 361
DB 61 PWLEREIEETTKLGFKHPVIG 82

RESULT 13
ABB19422
ID ABB19422 standard; Protein; 82 AA.
XX
AC ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
RESULT 12
ABB33985
ID ABB33985 standard; Peptide; 82 AA.
XX
AC ABB33985;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #1491 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
XX
OS WO200157277-A2.
XX
PN 09-AUG-2001.
XX
PD
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
DR
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX
PS Claim 27; SEQ ID NO 26620; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 82 AA;
Query Match 14.0%; Score 431; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-31;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 280 EVKDKNQVVELPIVDSLHPRPPYLPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
DB 1 EVKDKNQVVELPIVDSLHPRPPYLPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60

QY 340 PWLEREIEETTKLGFKHPVIG 361
DB 61 PWLEREIEETTKLGFKHPVIG 82
```

XX DE Protein #1421 encoded by probe for measuring heart cell gene expression.  
XX DE Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX OS Homo sapiens.  
XX PN WO200157274-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00666.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human hearts -  
XX Claim 15; SEQ ID NO 21192; 530pp; English.  
XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see AB21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 82 AA;  
Query Match 14.0%; Score 431; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-31;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 280 EVKDKNVQVVELPIVDSLHPRPPVPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVVELPIVDSLHPRPPVPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60  
QY 340 PWLEIEETTKKLGFKHPVIG 361  
DB 61 PWLEIEEATKKLGFKHPVIG 82  
RESULT 14  
AAM54747  
ID AAM54747 standard; Protein; 82 AA.  
XX AAM54747;  
XX 05-NOV-2001 (first entry)  
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 26852.  
XX Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human brains -  
XX Example 4; SEQ ID NO: 26852; 650pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.  
XX SQ Sequence 82 AA;  
Query Match 14.0%; Score 431; DB 22; Length 82;  
Best Local Similarity 96.3%; Pred. No. 2e-31;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 280 EVKDKNVQVVELPIVDSLHPRPPVPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339  
DB 1 EVKDKNVQVVELPIVDSLHPRPPVPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60  
QY 340 PWLEIEETTKKLGFKHPVIG 361  
DB 61 PWLEIEEATKKLGFKHPVIG 82  
RESULT 15  
AAM67139  
ID AAM67139 standard; Protein; 82 AA.  
XX AAM67139;  
XX 06-NOV-2001 (first entry)  
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 27445.  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00668.

```
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 27445; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 431; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-31; Mismatches 1; Indels 0; Gaps 0;
Matches 79; Conservative 2;
QY 280 EVKDKNVQVVELPIVDSLHPRPPYLPVLPEDLADRLRVHGDPAVWWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYLPVLPEDLADRLRVHGDPAVWWSQFVKYLIRPQ 60
QY 340 PWLEREIEETTKKLGFKHPVIG 361
Db 61 PWLEREIEETTKKLGFKHPVIG 82
```

Search completed: February 2, 2004, 08:41:46  
Job time : 40.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 15.5 Seconds  
(without alignments)  
1569.596 Million cell updates/sec

Title: US-09-971-773-23

Perfect score: 3081

Sequence: 1 MRATGSGWRIMLILFWAGT.....YKVRKIETVKYPTYPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2984	96.9	575	3	US-08-913-805A-10
2	2984	96.9	575	3	US-09-442-629-10
3	2885	93.6	575	3	US-08-913-805A-2
4	2885	93.6	575	3	US-09-442-629-2
5	131	4.3	25	3	US-08-913-805A-12
6	131	4.3	25	3	US-09-442-629-12
7	118.5	3.8	976	3	US-09-104-324B-4
8	116	3.8	214	5	PCT-US96-08950-2
9	116	3.8	214	5	PCT-US96-09127-2
10	116	3.8	229	4	US-08-630-915A-221
11	115	3.7	1027	4	US-09-914-259-27
12	115	3.7	3248	1	US-08-353-700-1
13	115	3.7	3248	5	PCT-US95-16216-1
14	111.5	3.6	644	4	US-09-198-452A-63
15	111	3.6	933	3	US-09-141-206-2
16	108.5	3.5	2482	1	US-08-328-254-6
17	108	3.5	418	4	US-09-198-452A-1056
18	108	3.5	933	3	US-09-107-149-17
19	107	3.5	922	3	US-09-141-206-6
20	107	3.5	933	3	US-09-107-149-2
21	106.5	3.5	636	4	US-09-198-452A-489
22	106.5	3.5	741	2	US-08-462-481-2
23	106.5	3.5	741	2	US-08-436-771-2
24	106.5	3.5	741	2	US-08-436-771-4
25	106.5	3.5	741	2	US-08-434-998-2
26	106.5	3.5	741	2	US-08-434-998-4
27	106.5	3.5	741	2	US-08-487-797-2

28 106.5 3.5 741 2 US-08-487-797-4  
29 106.5 3.5 741 2 US-08-701-005A-2  
30 106.5 3.5 741 2 US-08-479-895-2  
31 106.5 3.5 741 5 PCT-US95-02058-2  
32 106.5 3.5 741 5 PCT-US95-02058-4  
33 106 3.4 26 3 US-08-913-805A-3  
34 106 3.4 26 3 US-09-442-629-3  
35 106 3.4 251 4 US-08-630-915A-8  
36 106 3.4 1033 4 US-09-914-259-26  
37 105.5 3.4 1010 4 US-09-134-001C-5178  
38 105.5 3.4 1036 3 US-08-968-752B-4  
39 105.5 3.4 1036 4 US-09-536-224-4  
40 105 3.4 1230 2 US-08-968-542C-35  
41 104.5 3.4 973 3 US-09-107-149-19  
42 104.5 3.4 1074 2 US-08-768-147B-2  
43 104.5 3.4 1074 3 US-08-968-752B-2  
44 104.5 3.4 1074 3 US-09-107-149-3  
45 104.5 3.4 1074 4 US-09-536-224-2

#### ALIGNMENTS

RESULT 1  
US-08-913-805A-10  
; Sequence 10, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

Sequence 4, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 8, Appli  
Sequence 3, Appli  
Sequence 5178, Ap  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 35, Appli  
Sequence 19, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli

```
; MOLECULE TYPE: peptide
US-08-913-805A-10

Query Match
Best Local Similarity 96.2%; Score 2984; DB 3; Length 575;
Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSSRELKILAKLERLKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSSRELKILAKLERLKQONEDL 60

QY 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAEQIENYKQARNDLKGKHEILRRRIE 120
DB 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAEQIENYKQARNDLKGKHEILRRRIE 120

QY 121 NGAKELWFFLOSELKILKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180
DB 121 NGAKELWFFLOSELKILKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180

QY 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYGCQLHHVVCYCFMAYGTQRT 240
DB 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYGCQLHHVVCYCFMAYGTQRT 240

QY 241 LILESQNRWYATGWTVPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGWTVPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLVRVHGDPVWVVSQFVKYLIRPOPWLEREIETTKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLVRVHGDPVWVVSQFVKYLIRPOPWLEREIETTKLGFKHPVI 360

QY 361 GVHVRTDKVGTAAAFPIIEYVMVVEHFQLLARRMQVVKRYLATDDPSLLKEATK 420
DB 361 GVHVRTDKVGTAAAFPIIEYVMVVEHFQLLARRMQVVKRYLATDDPSLLKEATK 420

QY 421 YSNYEFISDINSISWAGLHNRRTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWAGLHNRRTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480

QY 481 TLHPDASANFSLDDIYFGGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHWGYS 540
DB 481 TLHPDASANFSLDDIYFGGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHWGYS 540

QY 541 KGVNRKLGRTGLYPSYKVKREKTIETVKYPTYPEAK 575
DB 541 KGVNRKLGRTGLYPSYKVKREKTIETVKYPTYPEAK 575

RESULT 2
US-09-442-629-10
; Sequence 10, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; UOZUMI, Naofumi
; SHIBA, Tetsuo
; YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA: US/09/442,629
; APPLICATION NUMBER: US/09/442,629
```

```
; FILING DATE: 18-No. 6291219-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
; FILING DATE: 23 JAN 1997
; APPLICATION NUMBER: JP 192260
; FILING DATE: 22 JUL 1996
; APPLICATION NUMBER: JP 162813
; FILING DATE: 24 JUN 1996
; APPLICATION NUMBER: JP 161648
; FILING DATE: 21 JUN 1996
; APPLICATION NUMBER: JP 10365
; FILING DATE: 24 JAN 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2356/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-442-629-10

Query Match 96.9%; Score 2984; DB 3; Length 575;
Best Local Similarity 96.2%; Pred. No. 6.1e-270;
Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSSRELKILAKLERLKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLFLFYGGHLVRNDHPDHSSRELKILAKLERLKQONEDL 60

QY 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAEQIENYKQARNDLKGKHEILRRRIE 120
DB 61 RMAESLRIPGPIQOGTATGRVRLVEQLVKAEQIENYKQARNDLKGKHEILRRRIE 120

QY 121 NGAKELWFFLOSELKILKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180
DB 121 NGAKELWFFLOSELKILKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGWR 180

QY 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYGCQLHHVVCYCFMAYGTQRT 240
DB 181 EKEADLTTELVRITTYLQNPDKCSKARKLVNKGCGYGCQLHHVVCYCFMAYGTQRT 240

QY 241 LILESQNRWYATGWTVPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGWTVPVSETCTDRSGISTGHSGEVKDKNVQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLVRVHGDPVWVVSQFVKYLIRPOPWLEREIETTKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLVRVHGDPVWVVSQFVKYLIRPOPWLEREIETTKLGFKHPVI 360

QY 361 GVHVRTDKVGTAAAFPIIEYVMVVEHFQLLARRMQVVKRYLATDDPSLLKEATK 420
DB 361 GVHVRTDKVGTAAAFPIIEYVMVVEHFQLLARRMQVVKRYLATDDPSLLKEATK 420

QY 421 YSNYEFISDINSISWAGLHNRRTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWAGLHNRRTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480

QY 481 TLHPDASANFSLDDIYFGGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHWGYS 540
DB 481 TLHPDASANFSLDDIYFGGQNAHQIAVYPHQRTKEEIPMEPGDIIGVAGNHWGYS 540

QY 541 KGVNRKLGRTGLYPSYKVKREKTIETVKYPTYPEAK 575
DB 541 KGVNRKLGRTGLYPSYKVKREKTIETVKYPTYPEAK 575
```

## RESULT 3

US-08-913-805A-2  
; Sequence 2, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-913-805A-2

Query Match 93.6%; Score 2885; DB 3; Length 575;  
Best Local Similarity 92.9%; Pred. No. 1.1e-260;  
Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;  
QY 1 MRATGSGNRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60  
DB 1 MRPTGSGNRWIMLILFANGTLLFYIGGHLVRDNDHSDHSRSLSKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPGPIDQGTATGRVRLBEQLVAKQEIENYKKQARNDLGKHHEILRRRIE 120  
DB 61 RMAESLRIPGPIDQGPASGRVRLAEQFMKAKEQIENYKQTKNGPGKDKHEILRRRIE 120  
QY 121 NGAKELWFLQSELKKLKEGNELOQRADELILLDLGHHSRIMTDLYLSQTDGAGWR 180  
DB 121 NGAKELWFLQSELKKLKEGNELOQRADELILLDLGHHSRIMTDLYLSQTDGAGWR 180  
QY 181 EKEAKDLTELVRRTTYLQNPDKCSKAKLVCMNKGCGYGCQLHHVYVCFMIAYGRTQT 240

DB 181 EKEAKDLTELVRRTTYLQNPDKCSKAKLVCMNKGCGYGCQLHHVYVCFMIAYGRTQT 240  
QY 241 LILESQNRVYATGGWETVFRPVSCTCTDRSGLSTGHSWGEVKDNQVQWVELPIVDSLHPR 300  
DB 241 LALESHNRVYATGGWETVFRPVSCTCTDRSGSSTGHSWGEVKDNQVQWVELPIVDSVHPR 300  
QY 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQWLEREIBETTKLGFKHPVI 360  
DB 301 PPYLPLAVPEDLADFLVRVHGDPAVWVVSQFVKYLIRPQWLEREIBETTKLGFKHPVI 360  
QY 361 GVHVRRTDKVGTAAAPHPIEYVMVHVEEHFOLLERRMKVKRYVLLATDDPSLLKEATK 420  
DB 361 GVHVRRTDKVGAAPHPIEYTVHVEEDFOLLARRMQDKRYVLLATDDPSLLKEATK 420  
QY 421 YSNYEFTSDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
DB 421 YPSYEFISDNSISWSAGLHNRYTENSRLGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFSLDIIYFQGNNAHQIAVYPHQRTKEBIPMPBGDIIGVAGNHNWYGS 540  
DB 481 ALHPDASANFRSLDIIYFQGNNAHQIAVYPHQRTKEBIPMPBGDIIGVAGNHNWYGS 540  
QY 541 KGVNRKLGKTGLYPSYKVRKEIETVKYPTYPEAK 575  
DB 541 KGVNRKLGRTGLYPSYKVRKEIETVKYPTYPEADK 575

## RESULT 4

US-09-442-629-2  
; Sequence 2, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/442,629  
; FILING DATE: 18-NO. 6291219-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-442-629-2

Query Match          93.6%; Score 2885; DB 3; Length 575;
Best Local Similarity 92.9%; Pred. No. 1,1e-260;
Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MEAWTGSNRWIMLILFANGTLLFYGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
DB 1 MRPWTGSNRWIMLILFANGTLLFYGGHLVRDNDHSRSLSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPGEPIDQGTATGRVRLVEQLVKAKEQIENYKKQAKGKDEILRRRIE 120
DB 61 RMAESLRIPGEPIDQGPASGRVRALEQFMKAKEQIENYKKQAKGKDEILRRRIE 120
QY 121 NGAKELWFFLQSELKLLKLEGNELQORHADELILLDGHHERSIMTDLVYLSQTDGAGWR 180
DB 121 NGAKELWFFLQSELKLLKLEGNELQORHADELFLDGHHERSIMTDLVYLSQTDGAGWR 180
QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNKNKGGCGYQGLHHVYVCFMIAVGTQRT 240
DB 181 EKEAKDLTELVRRTIYLNQPKDCSKAKKLVNKNKGGCGYQGLHHVYVCFMIAVGTQRT 240
QY 241 LLESQNRVYATGCHETVPRPVSCTDRSGLSSTGHWSGEVXKQVQVVELPIVDSLHPR 300
DB 241 LALESHNWRVYATGCHETVPRPVSCTDRSGSSTGHWSGEVXKQVQVVELPIVDSLHPR 300
QY 301 PPVPLAVPEDLADRLRLVHGDPAPVWVSQFVKYLIRPQPLERIEETTKLGFKHPVI 360
DB 301 PPVPLAVPEDLADRLRLVHGDPAPVWVSQFVKYLIRPQPLERIEETTKLGFKHPVI 360
QY 361 GVHVRTDKVGTAAFPHPLEIEYMHVVEHFQLLERRMKYDKRKYVLAITDDPSLLKEATK 420
DB 361 GVHVRTDKVGAFAFPHPLEIEYTVHVEEDFQLLARRMQVDKRVYLAITDDPALLKEATK 420
QY 421 YSNYFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADPLVCTFSQVCRVAYEIMQ 480
DB 421 YPSYFISDNSISWSAGLHNRYTENSRLRGVILDIHFLSQADPLVCTFSQVCRVAYEIMQ 480
QY 481 TLHPDASANFSLDDIYYFGGNAHQIAVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540
DB 481 ALHPDASANFRSLDDIYYFGGNAHQIAIYHPQRTKEEIPMEPGDIIGVAGNHNWGY 540
QY 541 KGVNRKLGKTLGYPYKVKREKIEYKPYTPAEK 575
DB 541 KGVNRKLGRTGLYPSYKVKREKIEYKPYTPAEADK 575

RESULT 5
US-08-913-805A-12
; Sequence 12, Application US/08913805A
; Patent No. 6054304
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; PRIORITY INFORMATION:
; PRIORITY NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171

MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-913-805A-12

Query Match          4.3%; Score 131; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 1,3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 KLGFKHPVIGVHVVRTDKVGTAAAF 376
DB 1 KLGFKHPVIGVHVVRTDKVGTAAAF 25

RESULT 6
US-09-442-629-12
; Sequence 12, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: UOZUMI, Naofumi
; APPLICANT: SHIBA, Tetsuo
; APPLICANT: YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,629
; FILING DATE: 18-No. 6291219-1999
; PRIORITY INFORMATION:
; PRIORITY NUMBER: US/08/913,805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
```



;; FILING DATE: 23 JAN 1997  
;; APPLICATION NUMBER: JP 192260  
;; FILING DATE: 22 JUL 1996  
;; APPLICATION NUMBER: JP 162813  
;; FILING DATE: 24 JUN 1996  
;; APPLICATION NUMBER: JP 161648  
;; FILING DATE: 21 JUN 1996  
;; APPLICATION NUMBER: JP 10365  
;; FILING DATE: 24 JAN 1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Toffenetti, Judith L.  
;; REGISTRATION NUMBER: 39,048  
;; REFERENCE/DOCKET NUMBER: 2356/3  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-429-1776  
;; TELEFAX: 202-429-0796  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-442-629-12  
  
Query Match 4.3%; Score 131; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 352 KLGFKHPVIGVHVRTDKVGTAAAF 376  
Db 1 KLGFKHPVIGVHVRTDKVGTAAAF 25  
  
RESULT 7  
US-09-104-324B-4  
; Sequence 4, Application US/09104324B  
; Patent No. 6232460  
; GENERAL INFORMATION:  
; APPLICANT: T reed, Ozlem; Sahin, Ugur; Pfreundschuh, Michael  
; TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,  
; TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of  
; TITLE OF INVENTION: No. 6232460mal Cells  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski LLP  
; STREET: 666 Fifth Avenue  
; CITY: New York City  
; STATE: New York  
; ZIP: 10103  
; COMPUTER READABLE FORM: Diskette, 3.5 inch, 144 kb storage  
; MEDIUM TYPE: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/104,324B  
; FILING DATE: 25-June-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/892,702  
; FILING DATE: 15-July-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 6232460man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5491  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 318-3000  
; TELEFAX: (212) 752-5958  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 976 amino acids  
; TYPE: amino acid

;; TOPOLOGY: linear  
US-09-104-324B-4  
  
Query Match 3.8%; Score 118.5; DB 3; Length 976;  
Best Local Similarity 20.5%; Pred. No. 0.054;  
Matches 90; Conservative 74; Mismatches 164; Indels 111; Gaps 19;  
  
QY 42 ELSKILAKLERLKQKQEDRRMAESLRIPEGPIDQGTATGR---VRVLEQLVKAKEQIE 98  
Db 435 ELKKVIGKEKTLVYENKQFEXIAEELKQTEQEL-IGLLQAREKEVHDLEIQLTAITTSQ 493  
  
QY 99 NYKQARNDLGKHILRRRIENGAKELWPFLOSELKKLKEGNELOHRADEILLDLGH 158  
Db 494 YYSKEVKD-----LKTELEN-EKLNKLTSHCNKL-SLENKELTQETSDMTLELKN 543  
  
QY 159 HERSI-----MTDLYVLSOTD-----GAGSWREKEAKDLTELVRRTTYLQNPKDC 204  
Db 544 QEDINNNKQOEERMLKQIENLQETETQLRNELEYVREELQKQDEVKCKLD--KSEENC 601  
  
QY 205 SKARKLVGNIN-----KCGYGCQHLHVYVCFMAYGTQRTLLILESQNRW 249  
Db 602 NNLRKQVENKNKIYIELQENKALKKGTAKQLN--VYEIKV---NKLELELESQAKQ 656  
  
QY 250 YATGGWETVFRPVSHCTDRSLSTGHWSGEVKQNVQVVELPIVDSLHPPPLPLAVP 309  
Db 657 F-----GEITDTYQKEIEDKKIS-----E 675  
  
QY 310 EDLADRL--LRVHGUPAV-----WVVSQFVKYLIRPQWLEREIEBETTKLGF-F 355  
Db 676 ENLLEEVEKAKVIADEAVKLOKEIDKRCQKHAEMVALMEKHQYDKIIBERDESLGLY 735  
  
QY 356 KHPVIGVHVRTDKVGTAAAFHPHIEYVMVHVEEHFOLLEREMKVDKRVLYLATDPPSLK 415  
Db 736 KSK---EQEQSSLRASLELSNLKALLSVKQLE-IEREKEKLEKE--AKENTATLK 789  
  
QY 416 EAKTKSYNEFISDHSISW 434  
Db 790 EKDKKTKTQTFLETPPIYW 808  
  
RESULT 8  
PCT-US96-08950-2  
; Sequence 2, Application PC/TUS9608950  
; GENERAL INFORMATION:  
; APPLICANT: OSTEOA INC.  
; TITLE OF INVENTION: Osteoclast Growth Regulatory Factor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM: Floppy disk  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/08950  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul C. Steinhart  
; REGISTRATION NUMBER: 30,806  
; REFERENCE/DOCKET NUMBER: FP-ND 2121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US96-08950-2

Query Match 3.8%; Score 116; DB 5; Length 214;

Best Local Similarity 38.0%; Pred. No. 0.009;

Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;

QY 509 AVYHPQRTKEEIPMEPGDIIIGVAG----NHNWYSGKGVNRKLGKGTGLYPSYKVKREKJET 564

Db 19 ALYTFEPTDLYFEGDIIITDMSDINWKGTSK-----GRTGLIPSNYVAEQAES 72

QY 565 VKYPTYPEAK 575

Db 73 IDNPLHEAAKR 83

#### RESULT 9

PCT-US96-09127-2

; Sequence 2, Application PC/TUS9609127

; GENERAL INFORMATION:

; APPLICANT: OSTEOA INC.

; TITLE OF INVENTION: Osteoclast Growth Regulatory Factor

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell & Flores LLP

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: La Jolla

; STATE: CA

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/09127

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Paul C. Steinhart

; REGISTRATION NUMBER: 30,806

; REFERENCE/DOCKET NUMBER: FP-ND 2122

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 214 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US96-09127-2

Query Match 3.8%; Score 116; DB 5; Length 214;

Best Local Similarity 38.0%; Pred. No. 0.009;

Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;

QY 509 AVYHPQRTKEEIPMEPGDIIIGVAG----NHNWYSGKGVNRKLGKGTGLYPSYKVKREKJET 564

Db 19 ALYTFEPTDLYFEGDIIITDMSDINWKGTSK-----GRTGLIPSNYVAEQAES 72

QY 565 VKYPTYPEAK 575

Db 73 IDNPLHEAAKR 83

#### RESULT 10

US-08-630-915A-221

; Sequence 221, Application US/08630915A

; Patent No. 6309820

#### GENERAL INFORMATION:

; APPLICANT: SPARKS, Andrew B.

; APPLICANT: HOFFMAN, No. 6309820h

; APPLICANT: KAY, Brian K.

; APPLICANT: FOWLKES, Dana M.

; APPLICANT: MCCONNELL, Stephen J.

; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

; TITLE OF INVENTION: USING SAME

; NUMBER OF SEQUENCES: 227

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/630,915A

; FILING DATE: 03-APR-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 1101-174

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 221:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 229 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-630-915A-221

Query Match

Best Local Similarity 38.0%; Score 116; DB 4; Length 229;

Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;

QY 509 AVYHPQRTKEEIPMEPGDIIIGVAG----NHNWYSGKGVNRKLGKGTGLYPSYKVKREKJET 564

Db 34 ALYTFEPTDLYFEGDIIITDMSDINWKGTSK-----GRTGLIPSNYVAEQAES 87

QY 565 VKYPTYPEAK 575

Db 88 IDNPLHEAAKR 98

#### RESULT 11

US-09-914-259-27

; Sequence 27, Application US/09914259

; Patent No. 6495336

; GENERAL INFORMATION:

; APPLICANT: Makowski, Lee

; APPLICANT: Hyman, Paul

; APPLICANT: Williams, Mark

; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

; FILE REFERENCE: 8471-010-999

; CURRENT APPLICATION NUMBER: US/09/914,259

; CURRENT FILING DATE: 2000-11-21

; NUMBER OF SEQ ID NOS: 180

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 27

; LENGTH: 1027

; TYPE: PRT



```
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match 3.7%; Score 115; DB 5; Length 3248;
Best Local Similarity 19.5%; Pred. No. 0.74;
Matches 110; Conservative 90; Mismatches 206; Indels 158; Gaps 28;

Qy 29 LVNDNDPHDSRELSTKILAKLERLKQONEDI---RMAESLRIPGPDIGTATGRVRV 85
Db 343 LVRTTAQYDQAS---TKYTALEQKLKLTDLSCORQNAESARCS----- 384
Qy 86 LEEQL-VKAKQIENYKQARNLCKDHEILRRRIE-----NGAKELWFFLQSELKLLK 139
Db 385 LEQKIKEKEPQBELSRQQRSFQILDQECIQMKARLTQELQQAQNMENVLQALDKLTS 444
Qy 140 LEGNELQRHADRIILLDLGHHERSIMTDLYLSQTDGAGWEKREKADLTTELVRRTIYLQ 199
Db 445 VK-QQLENNLEFFKQKLCRAEQAFQ-----ASQIKENELRRSMEEMKKNLLK 492
Qy 200 NPKDCSKARKLVNCINKGCGYGCQLHHVYCFMIAYGRTQRTILLESQNW----RYATGGM 255
Db 493 SHSE-QKARE-VCHLE-----AELKNIKQC-----LNQSNFAEMKAKNTSQ 533
Qy 256 ETVPFPVSETCTDRSLSTGHWSGEVKDKVQVVELP-----IVDSLHRRPPYLPLAVPE 310
Db 534 ETMLRDLQEKINQENSLT-----LEKLKVAADLEKQDCSQDLKKREHHI-----E 582
Qy 311 DLADRLLRVHGDPV-----WVWSQFVKYLIRPOPMLEREIE 347
Db 583 QLNDKLSKTERESKALLSALELKKEBELKEETLFCWKSSENEKLLTQ-----NESEKE 638
Qy 348 ETTKKLGKFPVIGVHVRTDKVGTAAFPHTPEEYVWVHEHFQLE---RRMKVDKRV 404
Db 639 NLQSKINHLETCL-----KTQOIKS-----HEY-----NERVRTLEMDRENLSVEIRNL 682
Qy 405 YLATDDPSLLKEAKTKSYNYEPIFSDNSISWSAGLNRYTEN-----SLRGVILDIHFLS 458
Db 683 HNVLSKSV--EVETQKLAYPELOQKA-EFSDQKHQKEIENMCLTSQLTGQVEDEHL 739
```

```
Qy 459 QADFLVCTFSQVCRVAYEIM--QTLHPDASANFHSLLDDIY-----YFGQNAHNQIATV 510
Db 740 QL-----LSNEIMDKRCYQDLHAYESLRDLKSKDASLVTWNEHQHRSLLA 786
Qy 511 YPHQRTKEEIPMEPFGDIIIGVAGN 534
Db 787 FDOQPAMHHSF----ANIIGEQS 806
```

## RESULT 14

US-09-198-452A-63

; Sequence 63 Application US/09198452A

; Patent No. 6559294

; GENERAL INFORMATION:

; APPLICANT: Griffsais, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; thereof and uses thereof, in particular for the diagnosis, prevention

; TITLE OF INVENTION: and treatment of infection

; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/09/198,452A

; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 63

; LENGTH: 644

; TYPE: PRT

; ORGANISM: Chlamydia pneumoniae

US-09-198-452A-63

Query Match 3.6%; Score 111.5; DB 4; Length 644;

Best Local Similarity 21.0%; Pred. No. 0.13; Mismatches 103; Conservative 73; Indels 125; Gaps 20;

Qy 43 LSKILAKLERLKQONEDLRMAESLRIPGPDIGDTA-----TCRVRLVEBQLVKAKDOI 97

Db 75 ISGFLLLRERREVSQVGLGPIGTG--IPVGPSAPSSSEIQKQKAKQILDLQLOELQDL 132

Qy 98 ENYKQARNDLG--KDHEILRRRIENGAKG-----LWFFLQSELKLLK 140

Db 133 DTDIQHVLSCGLKGLKDLCKDRGLLKDAKEKLQVDFVWKDMMMEFVLEQQVMDQESRYL 192

Qy 141 EG--NELQRHADIILLDLGHHERSIMTDLYLSQTDGAGWEKREKADLTTELVRRTIYL 198

Db 193 EGLIHEVQSIHAKLFVDDVNIIRSHLGSQVLPSEDVKGELLKRFAPK---EVAR---FM 246

Qy 199 QNPDCSKARKLVNCINKGCGYGCQLHHVYCFMIAYGRTQRTILLESQNWRYA----- 251

Db 247 KVTRD---IRKIAMAFNKN-AYGAANK---AFDKAFGSLETCLYKSLTKSYRDTFCDYK 298

Qy 252 -----TCGMETVFRPVSETCTDRSLSTGHWSGEVKDKVQVVELPIVDSLHRR- 300

Db 299 RAKILPDENNARSARAEQRFREYKDHMEDLN--ETVFWVWKE--DGRIDIEVLTAVGWMPDRY 354

Qy 301 PPYLPLAVPED-----LADRLLRVHGDPVAVWWSQFVKYLIRPOPMLEREIEETTK--- 351

Db 355 PEHLILEKRDVMSHQLWEATMRVKEAVTYISVAR-VAPEDGSGQQKQKFKQETKERL 413

Qy 352 -----KLGFHPVIGVHVRRTD---KVGTEAAFPHEE----- 381

Db 414 RCLXDLRDOECHRAQERLEKLTALYPEVSVSVETERERKFNLEKAYGNLEERYQSVQD 473

Qy 382 -----YMWVHEHFQLLRRMKYDKKRYVLYLATDDPSLLKEAK 418

Db 474 QEDYWTQKNREABFRAGTKVRSMEVEAHLQILENLLDCYKRLSKAETFALGVEREA 533

Qy 419 TKYSNYEPIFD 429

Db 534 TEEIYTLISD 544

## RESULT 15

US-09-141-206-2

; Sequence 2, Application US/09141206



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:40:23 ; Search time 217 Seconds  
(without alignments)  
550.686 Million cell updates

Title: US-09-971-773-23  
 Perfect score: 3081  
 Sequence: 1 MRAWTGSRWIMLILFANGT.....YKVREKLETVKVPTYPEAEK 575

Scoring table: BLOSUM62  
Gapop 10.0 ; Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10
                  Listing first 45

```

```

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	%			
1	3081	100.0	575	11	US-09-971-773-23	Sequence 23, Appl
2	3024	98.1	575	11	US-09-971-773-24	Sequence 24, Appl
3	2973	96.5	575	9	US-09-839-136-10	Sequence 10, Appl
4	2978	93.4	575	9	US-09-839-136-2	Sequence 2, Appl
5	2639	85.7	515	15	US-10-106-698-4558	Sequence 4558, Ap
6	431	14.0	82	9	US-09-864-761-34780	Sequence 34780, A
7	425	13.8	81	9	US-09-864-761-46107	Sequence 46107, A
8	133	4.3	1959	12	US-10-028-2484-36	Sequence 36, Appl
9	132	4.3	485	12	US-10-028-2484-319	Sequence 319, Ap
10	128	4.2	678	12	US-10-104-047-3419	Sequence 3419, Ap
11	127.5	4.1	1959	12	US-10-369-493-5427	Sequence 5427, Ap
12	125	4.1	1961	12	US-10-028-2484-106	Sequence 106, App
13	124.5	4.0	1960	12	US-10-028-2484-103	Sequence 103, App
14	124.5	4.0	1960	12	US-10-236-031B-62	Sequence 62, Appl
15	124	4.0	405	12	US-10-028-2484-104	Sequence 104, App
16	124	4.0	460	12	US-10-094-749-3191	Sequence 3191, Ap

Sequence 273, App  
Sequence 317, App  
Sequence 12, Appl  
Sequence 12504, A  
Sequence 596, App  
Sequence 107, App  
Sequence 105, App  
Sequence 221, App  
Sequence 19, Appl  
Sequence 43, Appl  
Sequence 260, App  
Sequence 73, Appl  
Sequence 18106, A  
Sequence 27, Appl  
Sequence 116, App  
Sequence 6374, Ap  
Sequence 79, Appl  
Sequence 7029, Ap  
Sequence 7030, Ap  
Sequence 49, App  
Sequence 286, App  
Sequence 63, Appl  
Sequence 21643, A  
Sequence 22825, A  
Sequence 6931, Ap  
Sequence 3, Appl  
Sequence 5702, Ap  
Sequence 6, Appl  
Sequence 4, Appl  
Sequence 5, Appl

## ALIGNMENTS

## RESULT 1

```

US-09-971-773-23
; Sequence 23, Application US/09971773
; Publication No. US20030115614A1
; GENERAL INFORMATION:
; APPLICANT: Yutaka KANDA
; APPLICANT: Mitsuo SATOH
; APPLICANT: Kazuyasu KAMURA
; APPLICANT: Kazuhisa UCHIDA
; APPLICANT: Toyohide SHINKAWA
; APPLICANT: Naoko YAMANE
; APPLICANT: Motoo YAWASAKI
; APPLICANT: No. US20030115614A1uo HANAI
; TITLE OF INVENTION: ANTIEODY COMPOSITION
; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971.7
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: JP 2000-30852
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/268,926
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Cricetus griseus
US-09-971-773-23

```

	Query Match	100.0%;	Score 3081;	DB 11;	Length 575;
	Best Local Similarity	100.0%;	Pred. No. 1.4e-266;		
	Matches 575;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MRAWTGSRWIMLI	LFANGTLLFYIGHLVRDNDPHDSRELSKILAKLERLQKQNDL	60	
Db	1	MRAWTGSRWIMLI	LFANGTLLFYIGHLVRDNDPHDSRELSKILAKLERLQKQNDL	60	
QY	61	RMAESLRIPEGPIDQGTATGRVRLSEQLVKAKEQIENYKKQARNIDLGKQHEILRRRE	120		

```
Db 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120
Qy 121 NGAKELWFFLOSELKLLKLEGNELQRADEILLDLGHHERSINTDLYYLSQTDGAGSWR 180
Db 121 NGAKELWFFLOSELKLLKLEGNELQRADEILLDLGHHERSINTDLYYLSQTDGAGSWR 180
Qy 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240
Db 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240
Qy 241 LILESQNRWYATGWTETFRPVSETCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGWTETFRPVSETCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300
Qy 301 PPYLPLAVPEDLADRLLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360
Db 301 PPYLPLAVPEDLADRLLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360
Qy 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEATK 420
Db 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEATK 420
Qy 421 YSNTEFISDINSISWAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Db 421 YSNTEFISDINSISWAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Qy 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540
Db 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540
Qy 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575
Db 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575
```

## RESULT 2

```
US-09-971-773-24
; Sequence 24, Application US/09971773
; Publication No. US200301156141
; GENERAL INFORMATION:
; APPLICANT: Yutaka KANDA
; APPLICANT: Mitsuo SATOH
; APPLICANT: Kazuyasu NAKAMURA
; APPLICANT: Kazuhisa UCHIDA
; APPLICANT: Toyohide SHINKAWA
; APPLICANT: Naoto YAMANE
; APPLICANT: Motoo YAMASAKI
; APPLICANT: No. US200301156141aluo HANAI
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL
; FILE REFERENCE: 249-202
; CURRENT APPLICATION NUMBER: US/09/971,773
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: JP 2000-308526
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/268,926
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 24
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-971-773-24
```

```
Query Match 98.1%; Score 3024; DB 11; Length 575;
Best Local Similarity 97.6%; Pred. No. 1.8e-261;
Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MRATGWSRWIMLILFANGTLLFYIGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60
Db 1 MRATGWSRWIMLILFANGTLLFYIGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60
```

```
Qy 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120
Db 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120
Qy 121 NGAKELWFFLOSELKLLKLEGNELQRADEILLDLGHHERSINTDLYYLSQTDGAGSWR 180
Db 121 NGAKELWFFLOSELKLLKLEGNELQRADEILLDLGHHERSINTDLYYLSQTDGAGSWR 180
Qy 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240
Db 181 EKEAKDLTELVRRIITYLQNPDKCSKARKLVNINCKGCGYCOLHHVVCYCFMIAIYGTORT 240
Qy 241 LILESQNRWYATGWTETFRPVSETCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGWTETFRPVSETCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300
Qy 301 PPYLPLAVPEDLADRLLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360
Db 301 PPYLPLAVPEDLADRLLRVHGDPAVWVSQFVKYLIRPQWLERIEBETTKLGFKHPVI 360
Qy 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEATK 420
Db 361 GVHVVRTDKVGTAAAFHPIIEEYVMVHVEHFQLLERMKVDKRVYLATDDPSLLKEATK 420
Qy 421 YSNTEFISDINSISWAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Db 421 YSNTEFISDINSISWAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
Qy 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540
Db 481 TLHPDASANFHSLLDDIYFVGQNAHNOIAVYPHQPRTKEEIPMEPGDIIGVAGNHNGYS 540
Qy 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575
Db 541 KGVNRKLGKTGLYPSYKVKREKIEYKTYPTYPEAEK 575
```

## RESULT 3

```
US-09-839-136-10
; Sequence 10, Application US/09839136
; Patent No. US20020081694A1
; GENERAL INFORMATION:
; APPLICANT: Naoyuki TANIGUCHI et al.
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 10
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-136-10
```

```
Query Match 96.5%; Score 2973; DB 9; Length 575;
Best Local Similarity 95.8%; Pred. No. 6.4e-257;
Matches 551; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
```

```
Qy 1 MRATGWSRWIMLILFANGTLLFYIGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60
Db 1 MRATGWSRWIMLILFANGTLLFYIGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60
Qy 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120
Db 61 RMAESLRIPGPIDOGTATGRVRVLEBQVKAKEQIENYKQARNDLGKDHEILRRRIE 120
```



Qy 121 NGAKELWFFLOSELKLLKLEGNELQRAHDEILDLGHHSRIMTDLYLSTQDAGDWR 180  
Db 121 NGAKELWFFLOSELKLLKLEGNELQRAHDEILDLGHHSRIMTDLYLSTQDAGDWR 180  
Qy 181 EKEAKDLTELQVRITTYLQNPDKCSKARKLVNKNKGGCGQLHHVVCYCFMIAYGRT 240  
Db 181 EKEAKDLTELQVRITTYLQNPDKCSKARKLVNKNKGGCGQLHHVVCYCFMIAYGRT 240  
Qy 241 LILESQNRWYATGWTETFRPVSCTCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300  
Db 241 LILESQNRWYATGWTETFRPVSCTCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300  
Qy 301 PPYLPLAVPEDLADRLRLVRHGDPAVWVVSQFVKYLIRPQWLEREIETTKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLVRHGDPAVWVVSQFVKYLIRPQWLEKEIEBATKLGFKHPVI 360  
Qy 361 GVHVRTDKVGEAAFPPIEYVMVHEHFOLERRMKVKKRVYLATDDPSLLKEATK 420  
Db 361 GVHVRTDKVGEAAFPPIEYVMVHEHFOLERRMKVKKRVYLATDDPSLLKEATK 420  
Qy 421 YSNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YPNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFSLDDIYFVGQNAHQIAVYPHQPRTKEEIPMBPGDIIGVAGNHNWGS 540  
Db 481 TLHPDASANFSLDDIYFVGQNAHQIAVYAHQPRTADEIPMBPGDIIGVAGNHNWGS 540  
Qy 541 KGVNRKLGKTGLYPSYKVKREKIETVKYPTYPEAEK 575  
Db 541 KGVNRKLGRTGLYPSYKVKREKIETVKYPTYPEAEK 575

## RESULT 4

US-09-839-136-2  
; Sequence 2, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Pig  
US-09-839-136-2

Query Match 93.4%; Score 2878; DB 9; Length 575;  
Best Local Similarity 92.7%; Pred. No. 2e-248;  
Matches 533; Conservative 16; Mismatches 26; Indels 0; Gaps 0;

Qy 1 MRWTGSRWMTMLILFANGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLRLKQONEDL 60  
Db 1 MRWTGSRWMTMLILFANGTLLFYIGGHLVRDNDHSRSLSKILAKLRLKQONEDL 60  
Qy 61 RMMAESLRIPGPDIDQGTATGRVRLBEQLVKAKEQIENYKQARNDLGKHDEILRRRIE 120  
Db 61 RMAGSLRIPGPDIDQGPASGRVRALESQFMKAKEQIENYKQTKNGPGKHDEILRRRIE 120  
Qy 121 NGAKELWFFLOSELKLLKLEGNELQRAHDEILDLGHHSRIMTDLYLSTQDAGDWR 180  
Db 121 NGAKELWFFLOSELKLLKLEGNELQRAHDEILDLGHHSRIMTDLYLSTQDAGDWR 180

Qy 181 EKEAKDLTELQVRITTYLQNPDKCSKARKLVNKNKGGCGQLHHVVCYCFMIAYGRT 240  
Db 181 EKEAKDLTELQVRITTYLQNPDKCSKARKLVNKNKGGCGQLHHVVCYCFMIAYGRT 240  
Qy 241 LILESQNRWYATGWTETFRPVSCTCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300  
Db 241 LALESHNRWYATGWTETFRPVSCTCTDRSGSSTGHSGEVKDKNVQVVELPIVDSVHPR 300  
Qy 301 PPYLPLAVPEDLADRLRLVRHGDPAVWVVSQFVKYLIRPQWLEREIETTKLGFKHPVI 360  
Db 301 PPYLPLAVPEDLADRLVRHGDPAVWVVSQFVKYLIRPQWLEKEIEBATKLGFKHPVI 360  
Qy 361 GVHVRTDKVGEAAFPPIEYVMVHEHFOLERRMKVKKRVYLATDDPSLLKEATK 420  
Db 361 GVHVRTDKVGEAAFPPIEYTVHVEEDFOLLARRMQVKKRVYLATDDPALLKEATK 420  
Qy 421 YSNYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YPSYEFISDINSISWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFSLDDIYFVGQNAHQIAVYPHQPRTKEEIPMBPGDIIGVAGNHNWGS 540  
Db 481 ALHPDASANFSLDDIYFVGPNANQIAIYPHQPRTEGEIPMBPGDIIGVAGNHNWGS 540  
Qy 541 KGVNRKLGKTGLYPSYKVKREKIETVKYPTYPEAEK 575  
Db 541 KGVNRKLGRTGLYPSYKVKREKIETVKYPTYPEADK 575

## RESULT 5

US-10-106-698-4658  
; Sequence 4658, Application US/10106698  
; Publication No. US20030105690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4658  
; LENGTH: 515  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-106-698-4658

Query Match 85.7%; Score 2639; DB 15; Length 515;  
Best Local Similarity 95.7%; Pred. No. 4.1e-227;  
Matches 489; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 65 ESLRIPGPDIDQGTATGRVRLBEQLVKAKEQIENYKQARNDLGKHDEILRRRIENGAK 124  
Db 5 KSLRIPGPDIDQGPALGRVRLBEQLVKAKEQIENYKQTRNGLGKHDEILRRRIENGAK 64  
Qy 125 ELWFFLOSELKLLKLEGNELQRAHDEILDLGHHSRIMTDLYLSTQDAGDWRKEA 184  
Db 65 ELWFFLOSELKLLKLEGNELQRAHDEFLDLGHHSRIMTDLYLSTQDAGDWRKEA 124  
Qy 185 KDLTELQVRITTYLQNPDKCSKARKLVNKNKGGCGQLHHVVCYCFMIAYGRTTILE 244  
Db 125 KDLTELQVRITTYLQNPDKCSKARKLVNKNKGGCGQLHHVVCYCFMIAYGRTTILE 184  
Qy 245 SQNRWYATGWTETFRPVSCTCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPRPPL 304  
Db 185 SQNRWYATGWTETFRPVSCTCTDRSGTSTGHSGEVKDKNVQVVELPIVDSLHPRPPL 244

QY 305 PLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPOPWLERIEETTKLGLFKHPVIGVHV 364  
Db 245 PLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPOPWLERIEETTKLGLFKHPVIGVHV 304  
QY 365 RTDKVGTAAAPHTPEEYVHVVEEHFOLLEREMKVDKRVYLATDDPSLLKEAKTKYSNY 424  
Db 305 RTDKVGTAAAPHTPEEYVHVVEEHFOLLARMQDKRVYLATDDPSLLKEAKTKYPNY 364  
QY 425 EFISDINSWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQTLHP 484  
Db 365 EFISDINSWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQTLHP 424  
QY 485 DASANFHSLLDDIYFPGGNAHQAIAVPHQPTKEEIPMEPGDIIGVAGNHNWGSKGVN 544  
Db 425 DASANFHSLLDDIYFPGGNAHQAIAVPHQPTKEEIPMEPGDIIGVAGNHNWGSKGVN 484  
QY 545 RKLGTGLPSYKVKREKIEYKPYTPEAEK 575  
Db 485 RKLGTGLPSYKVKREKIEYKPYTPEAEK 515

RESULT 6  
US-09-864-761-34720  
; Sequence 34720, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 34720  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL109847.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6  
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.6  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9  
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.80e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: AW002060.1, EVALUE 7.00e-44  
US-09-864-761-34720

Query Match 14.0%; Score 431; DB 9; Length 82;  
Best Local Similarity 96.3%; Pred. No. 8e-31;  
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 280 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPQ 339  
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRLRVHGDPAVWVWSQFVKYLIRPQ 60  
QY 340 PWLERIEETTKLGLFKHPVIG 361  
Db 61 PWLERIEETTKLGLFKHPVIG 82  
RESULT 7  
US-09-864-761-46107  
; Sequence 46107, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46107
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109847.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.40e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW002060.1, EVALUE 5.00e-43
US-09-864-761-46107

Query Match 13.8%; Score 425; DB 9; Length 81;
Best Local Similarity 96.3%; Pred. No. 2.7e-30;
Matches 78; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 280 EVKKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRLRVGDPVAVVWSQFVKYLIRPQ 339
Db 1 EVKKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRLRVGDPVAVVWSQFVKYLIRPQ 60

Qy 340 PWLEIEIETTKLGFKEHPVI 360
Db 61 PWLEKEIEEATKUGFKHPVI 81

RESULT 8
US-10-028-248A-36
; Sequence 36, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zertusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
```

```
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 36
; LENGTH: 1959
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-028-248A-36

Query Match 4.3%; Score 133; DB 12; Length 1959;
Best Local Similarity 15.1%; Pred. No. 0.041;
Matches 105; Conservative 95; Mismatches 191; Indels 160; Gaps 25;

Qy 32 DNDHPDHSRSLKILAKLERLKQO---NEDLARMAESLRIPGPIDQGTATGVRVLE 87
Db 1054 EGDSTDLS-D-QIAELQAQIAELKMLAKKEBELQ--AALARVEBEAAQKNWALKKIRELE 1110

Qy 88 EQLVKAKEQIB-----NYKQARNDLGKHDEILRRRIENGAKELWFFLQSELKKLKLLE 141
Db 1111 SQISELQEDLESEASRNKAQKQKRDGEELEALKTELEDTLDSI--AAQQLSRKRQOE 1168

Qy 142 GNEIQRHADEITLLDLGHHSRIMTDLYLSQTDGAGWEKREKADLTVELVORRITYLQNP 201
Db 1169 VNILKKTLEE--EAKTHEAQIQ-----EMRQKHSQAVEELAEQ----- 1204

Qy 202 KDCSKARKLVNINKGCGYGCQLHHVYVCFMIAYGTQRTLLILESONWRYATGGHWETVRFP 261
Db 1205 --LEQTKRVKANLEK-----AKQTL-----ENER----- 1226

Qy 262 VSETCTDRSLTCHWSGEVKNV--QVVVELPIVDSLHPRPPYPLAVPEDLADRLRV 319
Db 1227 -GELANVKKVLLQGRSEHKKVKAQQLQVQKFNIGER-----VETELADKVTK- 1277

Qy 320 HGDPVAVVWSQFVKYLIRPQVLEIREIBETTKLGFKEHPVIGVHVRRTDKVGTAA--FH 377
Db 1278 -----LQVELDNTVGLLS-----QSDSKSKSLTKDFS 1304

Qy 378 PIEYVMVVEHFQLLERRMKVKKRYLATDDFSLKKEATKYNSYEFISDINSISWSAG 437
Db 1305 ALESQLODTQELLQ-----EENRQKLSLSTK-----LQ-----VEDEKNSFRBQ 1344

Qy 438 LHNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVC-RVAYEIMOTLHPDASANPHSLDDI 496
Db 1345 LEEBEAKHNLEKQIATLH-AQVADMKKMEVSVCLETAEEVKRKLQD-----LEGL 1396

Qy 497 YFQGGQNAHQIAVYPHQRTKEIIPMEPGDIIIGVAGNHMNGYSKGVNRKLGKTGLPSY 556
Db 1397 ----SQRHEEKVAAVDKLEKTKRLQOELDLL--VDLDHQRSACNLEKKQKK---PQOL 1448

Qy 557 KVREKIETVKY 567
Db 1449 LAEKEKTSISKY 1459

RESULT 9
US-10-104-047-3419
; Sequence 3419, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
```

; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3419  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3419

Query Match 4.3%; Score 132; DB 12; Length 485;

Best Local Similarity 20.5%; Pred. No. 0.0063;

Matches 93; Conservative 72; Mismatches 137; Indels 152; Gaps 23;

QY 45 KILAKLERLK-QONEDLRMAESLRIPGPIDQGTATGRVRLVLEBQVLAKEQIENKKQ 103  
DB 13 KVTYKAEKRLKQEEERLKEE-----EASLK-----YEKE 45  
QY 104 ARNDLGKDHETLRRRIENGAKELWFFLOSELKKLKKLEGNELQORHADEI---LLDLG-- 157  
DB 46 -----EMERLEIQRIEKEK-----HRLKADLERNEELELYLLERCFF 86  
QY 158 -----HHERSINTDLYLSQTDGAGE-----WREKEAKDLTELQOR----- 193  
DB 87 EAEKLEKQETKLSQWKHYIQCDGSPDPVSAQEMNTFSLWKEKTNETFEVIEKSKVVLN 146  
QY 194 -----RITVLOPKDCSRKARKLVNCINIKGCGYCOLHHVVCFMIAYGTFQRTLILESQNW 248  
DB 147 LIEKLEKFLILETP-PCDLOQKNIIQYQESI---LQLELH---LKFGVATEILKQ--- 196  
QY 249 RYATGGWETVFRPVSETCTDRSLSTGHWSGVKDKNVQVVELPIDVLSLHPRPPYPLAV 308  
DB 197 -----ASTLAD-----LDGNNKEVIKDN---VTLYVWANKKPNRHSRVF 236  
QY 309 PE-----DLADRLLRVHGD-----PAVWW-----VSQFVKYLIR-PQP 340  
DB 237 SETQIGFIPRILATSDIAVRLHLYTHYVSAHPVTPSPKEYTSVAVTELKDDVKNVEK 296  
QY 341 WLEREIEETTKLGFKHPVIGVHVHVRTDKVGTAAAFPHIEEYVMVHVEEHFOLL--ERRMK 398  
DB 297 AISKVEEESKO-----QERGSHLQEBEIKVEEQQDIEVVMSSAESESAIKCEREMK 351  
QY 399 VDKKRVLYATDDPSLLKKAETKYSNYEFISDMSI 432  
DB 352 VLSETVSA--QLLLVENSSE--KPDFFEDNVV 380

## RESULT 10

US-10-369-493-5427

; Sequence 5427, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 5427

; LENGTH: 678

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-10-369-493-5427

Query Match

Best Local Similarity 4.2%; Score 128; DB 12; Length 678;

Pred. No. 0.024;

Matches 91; Conservative 82; Mismatches 186; Indels 108; Gaps 19;  
QY 38 HSSRELSKILAKLERLKQONEDLRMAESLRIPGPIDQGTATGRVRLVLEBQVLAKEQI 97  
DB 255 HLNGDMSDIELEASKMEQADEMSLAEEV-----FNERRAKEBELQGLEIEV 302  
QY 98 ENYK-----KQARNDLGKDHETLRRR-TENGAKELWFFLOSELKKL-KLE 141  
DB 303 EEOQKLEAVTHAMPQPMKEYEDLKSEAKLLRERVVEMEAKN--EDLDDDRISKYEIEIR 360  
QY 142 GNEIQRHA--DEILLDGHHERSINTDLYLSQTDGAGEWREKEAKDLTELQORRITYL 198  
DB 361 SNPUKKKAIQIQTDLTKQEEKLMEDMQSALTEPA---WRDKMSNMKQNLNADLVVIE 417  
QY 199 QNPK----DCSKARKLVNCINIKGCGYCOLHHVVCFMIAYGTFQ-----RTLILES 245  
DB 418 KQKHTVKDQISLASEELHEVDSQGEAQIMAHHTKYLLDLSSKSTMLDDTTENTYVQQIVYQ 477  
QY 246 QNWTYATGGWETVFRPVSE-----TCTDRSGLS--TGHWSGVKDKNVQV--- 288  
DB 478 QDIEFSDAVVLIILRKISANLKKVNLQEDQITDLDERGLTLQTN-VDELKEMHVRLOBEL 536  
QY 289 -----VELPI---VDSLHPRPPYPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQP 340  
DB 537 ISIDMELALNEEDINLETEK-----KIDQELAGVGNVDS----- 574  
QY 341 WLEREIEETTKLGFKHPVIGVHVHVRTDK--VGTEAAAFPHIEEYVMVHVEEHFOLLERMK 398  
DB 575 GLRQLEERQKRLDEAPARSHQMLEANVASIRNELHSIPGYSQH-----KMLRERLE 629  
QY 399 VDKKRVLYATDDPSLLKKAETKYSNYEFISDMSISWSAGLHNRITYEN 445  
DB 630 AVEKRTAAKSLDMSLRKTE-----IDYEDIKTESIRLOEYNNMLLTN 672

## RESULT 11

US-10-028-248A-106

; Sequence 106, Application US/10028248A

; Publication No. US20030235882A1

; GENERAL INFORMATION:

; APPLICANT: Shinketsu, Richard

; APPLICANT: Patturajan, Meera

; APPLICANT: Vernet, Corine

; APPLICANT: Casman, Stacie

; APPLICANT: Malyankar, Uriel

; APPLICANT: Shenoy, Suresh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gangoli, Esha

; APPLICANT: Miller, Charles

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Smithson, Glennda

; APPLICANT: Zernusen, Bryan

; APPLICANT: Liu, Xiaohong

; APPLICANT: Colman, Steven

; APPLICANT: Tchernev, Velizar

; APPLICANT: Si, Jingehe

; APPLICANT: Edinger, Shlomit

; APPLICANT: Stone, David

; APPLICANT: Sciore, Paul

; APPLICANT: Millet, Isabelle

; APPLICANT: Rothenberg, Mark

; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods

; FILE OF INVENTION: Thereof

; FILE REFERENCE: 21402-222

; CURRENT APPLICATION NUMBER: US/10/028,248A

; CURRENT FILING DATE: 2001-12-19

; PRIOR APPLICATION NUMBER: 60/256619

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/262959

; PRIOR FILING DATE: 2001-01-19

;; PRIOR APPLICATION NUMBER: 60/272408  
;; PRIOR FILING DATE: 2001-02-28  
;; PRIOR APPLICATION NUMBER: 60/285189  
;; PRIOR FILING DATE: 2001-04-20  
;; PRIOR APPLICATION NUMBER: 60/308039  
;; PRIOR FILING DATE: 2001-07-26  
;; PRIOR APPLICATION NUMBER: 60/311266  
;; PRIOR FILING DATE: 2001-08-09  
;; NUMBER OF SEQ ID NOS: 211  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 106  
;; LENGTH: 1959  
;; TYPE: PRT  
;; ORGANISM: Gallus gallus  
US-10-028-248A-106

Query Match 4.1%; Score 127.5; DB 12; Length 1959;  
Best Local Similarity 19.8%; Pred. No. 0.13;  
Matches 110; Conservative 79; Mismatches 177; Indels 189; Gaps 22;  
QY 42 ELSKILAKLE-RLKQNDLGRMAESLRIPEGPIDQGTATGRVRLVLEQLVKAKEQIE-- 98  
DB 1066 ELQAIAELKQLSKKEBELQ--AALARVEEBAQAQNMALKKIRELESQITELQEDLESE 1123  
QY 99 ----NYKQARNDLGKHDEILRRRIEN-----GAKELWFFLQSELKKLKGNEQLQRH 148  
DB 1124 RASRNKAQKQKDLGEELEALKTELEDYLDSPAQQELRSKREQEVTVLVKXLEDAKTH 1183  
QY 149 ADEILLDLGHHERSIMTDLVYLSQTDGAGWEKEAKDLTELIVORRITYLQNPDCSKAR 208  
DB 1184 EAQI-----QEMRQKHSQAIEELAEQ-----LEQTK 1209.  
QY 209 KLVCNINKCGYGCQHLHVYCFMAYGTQRTLLILESQNWRYATGWTVPVPRVSETCTD 268  
DB 1210 RVKANLEKA-----KQALESER-----AELSN 1232  
QY 269 RSGSLTGHWSGVKDKNV--QVVELPIVDSLHPRPPYLPVAPEDLADRL--LRVHGDP 324  
DB 1233 VKVLQKGDAEHKKKKVDAQLOELQVKTETGER-----VKTELAERVNKLQVELDNV 1285  
QY 325 VWWSQ-----FVKYLIRPOPWLREI-----EETTKLGFHPVIGVHVRRTDKVGT 372  
DB 1286 TGLLNQSDSKSTK-LAKDFSALSESQLODTQELLQEBETRLKLSFS-----TKLKQT 1334  
QY 373 EAAPFPIEYVMVVEEHQLLERRMKVKRVLVLTDPSSLKKAETKYSNVEFSDNSI 432  
DB 1335 EDEKNALKEQLEEEBAERKNELEKQISVLQQQAV-----EARKKMD----- 1374  
QY 433 SWSAGLHNRYTENSRLRGVILDIHFLSQADFLVCTPSSQVCRVAVYEMOTLHPDASANFHS 492  
DB 1375 -----DGLGC-----LETAEEAKKKLQKD-----LES 1396  
QY 493 LDDIYFGQNAHQIAVYVPHQRTKEIPMEPGDIIGVAGNHWNMGYSGVNRKLGKTGL 552  
DB 1397 LTQRY-----EKKIAVDKLEKTKRLQLQELDD-IAVDLHDHQRTVSNLEKQKK--- 1445  
QY 553 YPSYKVRKIEVTYK 567  
DB 1446 FDQLAAEEKNISAKY 1460

RESULT 12  
US-10-028-248A-103  
; Sequence 103, Application US/10028248A  
; Publication No. US20030235882A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Vernet, Corine  
; APPLICANT: Casman, Stacie  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Spytek, Kimberly

;; APPLICANT: Gangolli, Esha  
;; APPLICANT: Miller, Charles  
;; APPLICANT: Boldog, Ferenc  
;; APPLICANT: Li, Li  
;; APPLICANT: Taupier Jr, Raymond J  
;; APPLICANT: Kekuda, Ramesh  
;; APPLICANT: Smithson, Glennda  
;; APPLICANT: Zerhusen, Bryan  
;; APPLICANT: Liu, Xiaohong  
;; APPLICANT: Colman, Stevan  
;; APPLICANT: Tchernev, Velizar  
;; APPLICANT: Si, Jingsheng  
;; APPLICANT: Edinger, Shlomit  
;; APPLICANT: Stone, David  
;; APPLICANT: Sciore, Paul  
;; APPLICANT: Millet, Isabelle  
;; APPLICANT: Rothenberg, Mark  
;; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods  
;; FILE REFERENCE: 21402-22;  
;; CURRENT APPLICATION NUMBER: US/10/028,248A  
;; PRIOR FILING DATE: 2001-12-19  
;; PRIOR APPLICATION NUMBER: 60/256619  
;; PRIOR FILING DATE: 2000-12-19  
;; PRIOR APPLICATION NUMBER: 60/262959  
;; PRIOR FILING DATE: 2001-01-19  
;; PRIOR APPLICATION NUMBER: 60/272408  
;; PRIOR FILING DATE: 2001-02-28  
;; PRIOR APPLICATION NUMBER: 60/285189  
;; PRIOR FILING DATE: 2001-04-20  
;; PRIOR APPLICATION NUMBER: 60/308039  
;; PRIOR FILING DATE: 2001-07-26  
;; PRIOR APPLICATION NUMBER: 60/311266  
;; PRIOR FILING DATE: 2001-08-09  
;; NUMBER OF SEQ ID NOS: 211  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 103  
;; LENGTH: 1961  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-028-248A-103

Query Match 4.1%; Score 125; DB 12; Length 1961;  
Best Local Similarity 19.2%; Pred. No. 0.21;  
Matches 106; Conservative 94; Mismatches 191; Indels 162; Gaps 26;  
QY 32 DNDHPDSSRELKILAKLERLKOQ----NEDLRMAESLRIPEGPIDQGTATGRVRLV 87  
DB 1054 EGDSTDLSQ-QIAELQAIKELKQAKKEBELQ--AALARVEEBAQAQNMALKKIRELE 1110  
QY 88 EOLVKAKEQIE-----NYKQARNDLGKHDEILRRRIENGAKELWFFLQSELKKLKL 141  
DB 1111 SQISELOQDLSEERASRNKAQKQKRDGLGEELEALKTELEDITLDST--AAQELRSKRQ 1168  
QY 142 GNEQLQHADEITLLDIGHHERSIMTDLVYLSQTDGAGWEKEAKDLTELIVORRITYLQNP 201  
DB 1169 VNILKXLEE--FAKTHEAQIQ-----EMRQKHSQAIEELAEQ----- 1204  
QY 202 KDCSKARKLVNINKGCGYGCQHLHVYCFMAYGTQRTLLILESQNWRYATGWTVPVPRP 261  
DB 1205 --LEQTKRVKANLEK-----AKQTL-----ENER----- 1226  
QY 262 VSETCTDRSGLSTGHWSGVKDKNV--QVVELPIVDSLHPRPPYLPVAPEDLADRLRV 319  
DB 1227 -GELANEVYKVLQGGDRSEHKKRKKVEAQLOELQVKNFNEGER-----VRTEADKVTK- 1277  
QY 320 HGDPVAVWVVSQFVKYLIRPOPWLREIEBETTKLGFHPVIGVHVRRTDKVGTAA--FH 377  
DB 1278 -----LQVELDNVTGLS-----QSDSKSKLTKDFS 1304  
QY 378 PEEYVMVVEEHQLLERRMKVKRVLVLTDDPSLLKKAETKYSNVEFSDNSISWAG 437  
DB 1305 ALESQLODTQELLQ-----BENRQKLSLSTK-----LKQ-----VEDEKNSFREQ 1344



```
Qy 88 EQLVKAQEIE-----NYKQARNDLGKDHIEILRRRIENGAKELWFFLQSELKKLKKLE 141
Db 1111 SQISELQEDLESERASRNKAQKQKRDLEGELEALKTELEDLTDST--AAQQLSKREQE 1168
Qy 142 GNEIQRHAEILLDLGHHHSIMTDLVYLSOTDAGWEKEAKDLTELVRRTYILQNP 201
Db 1169 VNILKKTLEE---EAKTHEAQIQ-----EMRQKHSQAVBELAQ-----1204
Qy 202 KDCSKARKLVNINKGCGYGCQLHHVVVCFMAYGTORTLILESONWRVATGGWETVFRP 261
Db 1205 --LEQTKRVKANLEK-----AKQTL-----ENER-----1226
Qy 262 VSETCTDRSGLSGTHWSGEVDKQNV--QVVELPIVDSLHPRPPYLPPLAVPDLADRLIRV 319
Db 1227 -GELANEVKKVLLQKGDSEHKKVKAQQLQELQVKNFEGE-----VRLADKVTK- 1277
Qy 320 HGDPAMVWVSFVKYLRPQWLEREIEETTKLGFHPVIGVHVRTDKVGTAA--FH 377
Db 1278 -----LOVELDNVTGLLS-----QSDSKSKSLTKDIFS 1304
Qy 378 PIEEYVHVHVEHFOLERRMKVDKRVVYLATDDPSLLKEAKTKYSNYEFISDINSISAG 437
Db 1305 ALESQLOTOBLLQ-----BENQKJSLSTK-----LQ-----VEDEKNSFREQ 1344
Qy 438 L-HNRYTENSIRGVILDIHFLSQADFLVCTFSSQVC-RVAYEIMQTLHPDASANFHSLDD 495
Db 1345 LEEEEAKGNLEKQIATLH-AQVADMKKMKMEDSVCCLETAEEVKRKLQD-----LEG 1396
Qy 496 IYFYGQNAHNOIAVYHPQRTKBEIPMEPGDIIIGVAGNHWGYSKGVNRKLGKGTGLYPS 555
Db 1397 L-----SQRHEEKVAAYDKLEKTKRLQELDLIL-VLDLHQRSQACNLEKKQKK---FDQ 1448
Qy 556 YKVREKIEYVY 567
Db 1449 LLAEKTIISAKY 1460
```

## RESULT 15

```
US-10-094-749-3191
; Sequence 3191, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIG, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3191
; LENGTH: 405
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-094-749-3191
```

```
Query Match 4.08; Score 124; DB 12; Length 405;
```

```
Best Local Similarity 21.9%; Pred. No. 0.025;
Matches 79; Conservative 54; Mismatches 109; Indels 118; Gaps 18;
```

```
Qy 42 ELSKILAKLERLKQONEDLRMAESLRIPGPIDQGTATGRVVRVLEEQLVKAQEIQENYK 101
Db 116 ELSKY---KEIINRQKAEIQNLDDKVK---TADQ-----LQELQKQKQEIENLK 159
Qy 102 KQAR-----NDLQKQHEILRRRIENGAKELWFFLQSELKKLKKL--EGNELQRHAEIL 153
Db 160 EEVESLNSLNDLQKDIEGSRKR---ESELLLFERLTSKNAQLQSESNLSQSFQDKVS 215
Qy 154 LDLGHHHSIMTDLVYLSOTD-----GAGWEKEAKDL-----187
Db 216 CS-----ESQLOSQCEQMKQTNINLESRLLEEELRKEEVQTLQAEACRQTEVKALSTQV 271
Qy 188 ---TELVRRTYILQNPKDCSK---ARKLVNINKGCGYGCQLHHVVVCFMAYGTQR 239
Db 272 EELKDELVTQRRKGIASSIKDLTKQLQARRKLDQVSSG-----SYDKE- 314
Qy 240 TLILESQNRVATYGTWETVFRPVSETCTDRSGLSGTHWSGEVKQKNVQVVELPIVDSLHP 299
Db 315 ---VSSMGSRSSSG---SLNARSSAEDRSPTGT-----SSVAVDNFPQVDK--- 356
Qy 300 RPPYLPPLAVPEDLADRLIRVHG--DPAVWVVSQFVKYLRPQWLEREIEETTKLGPKH 357
Db 357 -----AMLIERVRLRKAHARKNEKIEFMEDHIKQLV-----BEIRKTKKI-YKH 401
```

Search completed: February 2, 2004, 08:52:03

Job time : 219 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 17 Seconds  
(without alignments)  
3252.764 Million cell updates/sec

Title: US-09-971-773-23

Perfect score: 3081

Sequence: 1 MRAWTGSWRIMLILFAWG.....YKVRKTIETVKYTPYPAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2984	96.9	575	JC5432	glycoprotein 6- $\alpha$ p
2	995	32.3	818	T32154	hypothetical prote
3	133	4.3	1173	T43527	sp8 protein - f1ss
4	133	4.3	1727	T50073	myosin-like coiled
5	128	4.2	678	H88187	protein C18H9.8 [i
6	127.5	4.1	1959	A33977	myosin heavy chain
7	127	4.1	946	S28061	SCP1 protein - rat
8	125	4.1	1961	A61231	myosin heavy chain
9	124.5	4.0	845	I48176	synaptonemal compl
10	124	4.0	1025	H86250	hypothetical prote
11	124	4.0	2442	T08621	centrosome associa
12	121.5	3.9	3187	JC5837	364K Golgi complex
13	120.5	3.9	978	A70387	conserved hypothet
14	119.5	3.9	1270	T09194	adaptor protein in
15	119.5	3.9	2007	B43402	myosin heavy chain
16	119	3.9	1871	D96796	probable heat choc
17	118.5	3.8	1427	S22695	restin - human
18	118	3.8	581	S46327	gene cel IF A3 pro
19	118	3.8	1999	I S21801	myosin heavy chain
20	117.5	3.8	1282	J80120	glycoprotein A - m
21	116	3.8	993	S49461	synaptonemal compl
22	116	3.8	1156	B70356	chromosome assembl
23	115	3.7	1027	S37711	kinesin heavy chain
24	115	3.7	1976	A59252	myosin heavy chain
25	115	3.7	2326	T29140	hypothetical prote
26	114.5	3.7	503	A43336	hypothetical TPR/T
27	114.5	3.7	1392	A43336	microtubule-vesicl
28	114.5	3.7	2017	A36014	myosin heavy chain
29	114.5	3.7	2057	S61477	myosin II heavy ch

ALIGNMENTS

RESULT 1

JC5432

glycoprotein 6- $\alpha$ -L-fucosyltransferase (EC 2.4.1.68) precursor - human

N;Alternate names: alpha 1-6 FucT; GDP-L-fucose:N-acetyl-beta-D-glucosaminyl:6- $\alpha$ -L-f

C;Species: Homo sapiens (man)

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000

C;Accession: JC5432; PC4322

R;Yanagidani, S.; Uozumi, N.; Ihara, Y.; Miyoshi, E.; Yamaguchi, N.; Taniguchi, N.

J. Biochem. 121, 626-632, 1997

A;Title: Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha

A;Reference number: JC5432; MUID:97279058; PMID:9133635

A;Accession: JC5432

A;Molecule type: mRNA

A;Residues: 1-575 <YAN1>

A;Cross-references: DDBJ:D89289; NID:G2055306; PIDN:BAA19764.1; PID:G2055307

A;Accession: PC4322

A;Molecule type: protein

A;Residues: 68-87;352-376;419-432 <YAN2>

C;Comment: This enzyme catalyzes the transfer of fucose from GDP-fucopyranoside to aspar

C;Genetics:

A;Gene: GDB:FUT8

A;Cross-references: GDB:9786294; OMIM:602589

A;Map position: 14q23-14q23

C;Superfamily: human glycoprotein 6- $\alpha$ -L-fucosyltransferase

C;Keywords: glycosyltransferase; hexosyltransferase

F;1-22/Domain: signal sequence #status predicted <Sig>

F;23-575/Product: glycoprotein 6- $\alpha$ -L-fucosyltransferase #status predicted <MAT>

Query Match 96.9%; Score 2984; DB 1; Length 575;

Best Local Similarity 96.2%; Pred. No. 2.2e-186;

Matches 553; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MRAWTGSWRIMLILFAWGTLFFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60

Db 1 MRPWTGSWRIMLILFAWGTLFFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60

Qy 61 RMAESLRIPRGPIDQGTATGVRVLEBQLVKAQEIENYKQARNDLGKHHEILRRRIE 120

Db 61 RMAESLRIPRGPIDQGTATGVRVLEBQLVKAQEIENYKQARNDLGKHHEILRRRIE 120

Qy 121 NGAKELPFFLOSELKLLKLEGNELQRADEILLDLGHHSRIMTDLYLSQTDGAGBWR 180

Db 121 NGAKELPFFLOSELKLLKLEGNELQRADEILLDLGHHSRIMTDLYLSQTDGAGBWR 180

Qy 181 EKEAKDLTELQVRITTYLQNPDKCSKAKKLYCNINKKCGYCOLHHVYVCFMAYGTQRT 240

Db 181 EKEAKDLTELQVRITTYLQNPDKCSKAKKLYCNINKKCGYCOLHHVYVCFMAYGTQRT 240

Qy 241 LILSQNWRVATGGWETVFRPVSETCTDRSLSTGHSGEVKDKNVQVVELPIVDSLHPR 300

Db 241 LILSQNWRVATGGWETVFRPVSETCTDRSLSTGHSGEVKDKNVQVVELPIVDSLHPR 300

hypothetical prote  
hypothetical prote  
myosin heavy chain  
hypothetical prote  
hypothetical prote  
hypothetical prote  
myosin heavy chain  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypothet  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
citron - mouse

QY 301 PPYLPLAVPEDLADRLRVHGDPAVWVSQFVKYLIRPQWLERIEETTKKLGKHPVI 360  
DB 301 PPYLPLAVPEDLADRLRVHGDPAVWVSQFVKYLIRPQWLERIEETTKKLGKHPVI 360  
QY 361 GVHVRTDKVCTEAPHPFIEEYVHVEHQLERMKVVKVYVLTATDDPSLLKEATK 420  
DB 361 GVHVRTDKVCTEAPHPFIEEYVHVEHQLERMKVVKVYVLTATDDPSLLKEATK 420  
QY 421 YSNYEFISDINSISWSAGLHNRNTENSRLGVLDIHFLSOADFLVCTFSSQVCRVAYEIMQ 480  
DB 421 YNRYEFISDINSISWSAGLHNRNTENSRLGVLDIHFLSOADFLVCTFSSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFHSLLDIYFYGQNAHQAIVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540  
DB 481 TLHPDASANFHSLLDIYFYGQNAHQAIVYHPQRTKEEIPMEPGDIIGVAGNHNWGS 540  
QY 541 KGVNRKLGRTGLYPSYKVKREKIEETVKYPTYPEAK 575  
DB 541 KGVNRKLGRTGLYPSYKVKREKIEETVKYPTYPEAK 575

RESULT 2  
T32154  
hypochemical protein C10F3.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32154  
R:Davidson, S.; Wohlmann, P.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid C10F3.  
A:Reference number: Z21127  
A:Accession: T32154  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-818 <DV>  
A:Cross-references: EMBL:AF022968; PIDN:AAB69888.1; GSPDB:GN00023; CESP:C10F3.6  
A:Experimental source: strain Bristol N2; clone C10F3  
C:Genetics:  
A:Map position: 5  
A:Introns: 30/1; 72/2; 235/2; 258/3; 285/3; 315/3; 457/3; 491/1; 544/3; 556/1; 626/1; 64

Query Match 32.3%; Score 995; DB 2; Length 818;  
Best Local Similarity 36.7%; Pred. No. 7.4e-57;  
Matches 213; Conservative 116; Mismatches 188; Indels 64; Gaps 11;  
QY 3 AWTGSRWIMLLIPAWGTLIFYGHLVRDNDHPDHSRELSKILAKLERLKOQNEDLRR 62  
DB 6 AAVGTVMWMTFLYLSQL-----SNNQSGGDSIRAWRQTKEADIKLQEQNEDLKS 56  
QY 63 MAESLRIPEGPIDQGTATGRVRLVEQLVKAKEQIENYKQARN-----107  
DB 57 ILEK-----QERNDQHKINEQSHQPPNPENSLPKPVPKELISKP 101  
QY 108 --LGK-DHEILARRIENGAKELWFFLOSELKKLKEGNEIQRHADEILLDGHHERSIM 164  
DB 102 SILGPVQEQVQRMLDRIEMFYLLHSQTIE-----NSTKILLET--QMISIM 148  
QY 165 TDLYLSQTDGAGEREKADLTTELVRITYLONPKDCSKARKLVNINKNGCGYCOL 224  
DB 149 GLSAQLEKLEGESEERFPQRTAITQIRIFKSEKLQNPACSEAKTLVNCNLKDECGFGCOL 208  
QY 225 HHVYCFMAYGTQRTLL--ESQNRVATGWETVFRPVSETCTDRSGLSGTGHSWGE--280  
DB 209 HHVYCAITAFATQRMVVLKRDSSWKYSSHGWTSPFKLLSKCSFDE--AVGTEAKFP 265  
QY 281 VKDKVQVVELPIVDSLHPRPYLPVLADELLRVHGDPAVWVSQFVKYLIRPQP 340  
DB 266 AEPSPARVSLGIVDSLITKPTFLPQAVPEQLLESILTSLSHPPAPFFVGTFTSYLMRFNS 325  
QY 341 WLERIEETTKLGF-KHPVIGVHVRTDKVCTEAPHPFIEEYVHVEHQLERMKVVKVYVLTATDDPSLLKEATK 398  
DB 326 ATQEKLDKALKSIPDKGPIVLGQIRRTDKVGTAAFAHALKEYMEWTIWFVKESKROGK 385

QY 399 VDKRVVLTATDDPSLLKEATKYSNYEFISDINSISWSAGLHNRNTENSRLGVLDIHFLS 458  
DB 386 PLERRIFIASDDPTVWPEAKNDYENYEVYGSTEIAKTAQLNRYTDSALMGVITDIYLS 445  
QY 459 QADFLVCTFSSQVCRVAYEIMQTLHPDASANFHSLLDIYFYGQNAHQAIVYHPQRTK 518  
DB 446 KNYLVCTFSSQVCRMGYELRQPSGADGSKFHSLLDIYFYGQQAHEVIVIEDHIAQNN 505  
QY 519 BEIPMEPGDIIGVAGNHNWGSKGVNRKLGRTGLYPSYKVR 559  
DB 506 KEIDLKVGDKVGIAGNHNWGSKGTNRQTYKEGVFPYSKIR 546

RESULT 3  
T43527  
sp8 protein - fission yeast (Schizosaccharomyces pombe) (fragment)  
C:Species: Schizosaccharomyces pombe  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Feb-2000  
C:Accession: T43527  
R:Jimenez, M.; Goday, C.  
submitted to the EMBL Data Library, June 1997  
A:Description: Schizosaccharomyces pombe myosin-like.  
A:Reference number: 222545  
A:Accession: T43527  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1173 <JIM>  
A:Cross-references: EMBL:AF010473; PIDN:AAB65416.1  
A:Experimental source: strain 972h (-)  
C:Genetics:  
A:Gene: sp8

Query Match 4.3%; Score 133; DB 2; Length 1173;  
Best Local Similarity 21.2%; Pred. No. 1.3;  
Matches 110; Conservative 74; Mismatches 226; Indels 108; Gaps 20;  
QY 34 DHPDHSRELSKILAKLERLKOQNEDLRRMAESLRIPEGPIDQGTATGRVRLVEQLVKA 93  
DB 612 NHVKRISKEMBIISQORQLPLENKLRKTVESSNRVIADLQRG-----ITEKDVST 664  
QY 94 KEQI---ENY-----KKQARNDLGKDHEI---LRRRIENGAKELWFF-----129  
DB 665 SESYGRSNLYNMVALLNESKSLRENLERNEEVITELREKIEITLKTDLANFLNKEOLE 724  
QY 130 --LOSELKKLKEGNE--ELQRHADEILLDGHHERSIMTDLYLSQTDGAGEREKAK 185  
DB 725 SLOTEKAAVKKLENSNEEYKRNQEIILSL-NSSTSTSSDASRLK-----NELVSK--776  
QY 186 DLTELVRITYLONPKDCSKARKLVNINKNGCGYCOLHHVYCFMAYGTQRTLLS 245  
DB 777 NLIELNQEIHLKSELETVKSKSEDLNER-----AQNQSKIEQLEL 819  
QY 246 QNRVYATGWETVFRPVSETCTDRSGLSGTGHSWGEKDKNVQVVELPIVDSLHPRPYLP 305  
DB 820 KNTKLA--AAMRTKYEVVNVKSLKHNQIRQOLSKTSELEAKVAECHOQLNQLNKPSTP 878  
QY 306 LAVPEDLADRLRVHGDPAVWVSQFVKYLIRPQWLERIEETTKKLGKHPVIGVHVR 365  
DB 879 TATTQ-----SEPTSVSLEEFNS-----TKEELSSTQKLEIMDLNTTKE 920  
QY 366 RTDKV-----GTEAAFH-PIEYM-----VHVEHQLERMKVVKVYVLTATDDPS 412  
DB 921 ELEKVRQNSKSEGTSDTEIPNEEEMERKVKMQEVLRLSRISAKELQKELLARKQV 980  
QY 413 LLKEAK---TKYSNYEFISDINSISWSAGLHNRNTENSRLGVLDIH-----LSQA 460  
DB 981 LQDQVKALQETVWSEEAESASVHADTKOLENLKKTTEMLSVTFQVIFNESIDFSTSTA 1040  
QY 461 DFLVCTFSSQVCRVAYEIMQTLHPDASANFH--SLDDI 496  
DB 1041 DF--TTFQKEWEKEREILQKDVVEQVAQSHQKQLDNI 1076

RESULT 4  
T50073  
myosin-like coiled-coil protein sp8 [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: T50073  
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25034  
A:Accession: T50073  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1727 <MCD>  
A:Cross-references: EMBL:AL133357; PIDN:CAB62414.1; GSPDB:GN00066; SPDB:SPAC1486.04C  
A:Experimental source: strain 972h(-); cosmid c1486  
C:Genetics:  
A:Gene: SPDB:SPAC1486.04C  
A:Map position: 1

Query Match 4.3%; Score 133; DB 2; Length 1727;  
Best Local Similarity 21.2%; Pred. No. 2.3; Mismatches 226; Indels 108; Gaps 20;  
Matches 110; Conservative 74

Qy 34 DHPDSSRELKILAKLERLQKQEDLRMAESRIPEGPIDQGTATGRVRVLEQLVKA 93  
Db 1166 NNVKRIKEMELISQORQLLFLENKKLRTVSSNRVIADLQRG-----ITEKDVST 1218  
Qy 94 KEQI---ENY-----KKQANDLGKDHETLRRRIENGAKELWFF----- 129  
Db 1219 SESVGSNLYNMVALLNESKSLRENLERNEEVITELREKIEITLKTDLANFRLNKEQLE 1278  
Qy 130 --LQSELKKLKEGN--ELORHADEIILLGHHSIRMTDLYLSQTDGAGWREKEAK 185  
Db 1279 SLOQTEKAAVKKLENSBEYKRHNQEIILLSL--NSSTSTSSDASRLK-----NELVSK 1330  
Qy 186 DLTELVRRIITVLPQKDCSKARKLVNINKGCGYGCQLHHVYCFMIAVGTQRTILLES 245  
Db 1331 NLIEELNQEIGHKSELTFTVSKSEDLNER-----AQNSKIEQLEL 1373  
Qy 246 QNWRYATGQWTFVFPVSECTCDRSLSTGHWSEGVKQNVQVVELPIVDSLHPRPPYLP 305  
Db 1374 KNTKLA-AAWRTKYEQVQVNSLEKHQIQRLSQKTSLEAKVAECHOLNEQLNKPSTP 1432  
Qy 306 LAVPEDLADRLRVHGDPAVWVVSQVFKYLTPQPLWLEREIEETTKLGFKHPVIGVHR 365  
Db 1433 TATTO-----SEFSTVTSLEEFNS-----TKEELSSQTKLSLIMDLTNTTKE 1474  
Qy 366 RTDKV-----GTEAAPH-PIEEM-----VHVEEHFOLLERRMKVDRVYLATDDPS 412  
Db 1475 ELEKVRQNSKSEGTSTKDTETPNEEMERKVKVQOEVLRLSRINAKELQKELLRKQNOV 1534  
Qy 413 LLKEAK----TKYSNYEFISDNSISWSAGLHNRYTENSRLRGVILDIHF-----LSQA 460  
Db 1535 LQDQVKALQETVWSSEAEASVHADTKDLENLKKTEEMLSVTFQVFNESISDFSTSTA 1594  
Qy 461 DPLVCTFSQVCRVAYEIMQTLHPDASANFH--SLDDI 496  
Db 1595 DF--TTFVQKWEKREILQKDVQVQAQSHQKQLDNI 1630

RESULT 5  
H88187  
protein C18H9.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: H88187  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_elegans/  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: H88187  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <STO>  
A:Cross-references: GB:chr\_XI; PID:g722387; GSPDB:GN00020; CBSP:C18H9.8  
C:Genetics:  
A:Gene: C18H9.8  
A:Map position: 2

Query Match 4.2%; Score 128; DB 2; Length 678;  
Best Local Similarity 19.5%; Pred. No. 1.4; Mismatches 82; Indels 108; Gaps 19;  
Matches 91; Conservative 82

Qy 38 HSSRELKILAKLERLQKQEDLRMAESRIPEGPIDQGTATGRVRVLEQLVKAQKQI 97  
Db 255 HLNGDMSDIELEASKMEQADEMSLAEV-----FNERRAKEELQGLEIEV 302  
Qy 98 ENYK-----KQANDLGKDHETLRR--IENGAKELWFFIQLSELKKLK-KLE 141  
Db 303 EEQKLNNAVTHAMPQMKKEYEDLKSEAKLLRERVVEMEAKN--EDLDDRISKEYEIR 360  
Qy 142 GNEIQRHA---DEILLDGHHSIRMTDLYLSQTDGAGWREKEAKDLTELVRRIYL 198  
Db 361 SNPLKKKAIQLQETLDTLKKOEKLMEDMQSALTPEA---WRDKMSNMKQLNADLVIE 417  
Qy 199 QNPK---DCSKARKLVNINKGCGYGCQLHHVYCFMIAVGTQ-----RTLILES 245  
Db 418 KQHTVQDQISLASEELHEVDSQEAQIMAHHTKYLLSKSTMLDDTTENYPOQIVYQ 477  
Qy 246 QNWRYATGQWTFVFPVSE-----TCTDRSGLS---TGHWSGEVKKQNVQV--- 288  
Db 478 QDIEFSDAVVILIRKISANLKVNLQEDITDLDERGLTLQTN-VDELKEMHVRLOBEL 536  
Qy 289 -----VELPI---VDSLHPRPPYLPVPELDALRLRVHGDPAVWVVSQVFKYLIRPQ 340  
Db 537 ISIDMELALNEETDNLTEK-----KIDQELAGVGNVDS----- 574  
Qy 341 WLERIEETTKLGFKHPVIGVHRRTDK--VGTEAAPHPIEEMVHVEEHFOLLERRMK 398  
Db 575 GLRQELERQKRLDEAPARSHQMLEANVASITNELHSIPGYSOH-----KMLRERLE 629  
Qy 399 VDKKRVYLATDDPSLLKEAKTKYSNYEFISDNSISWSAGLHNRYTEN 445  
Db 630 AVEKRTAAKSLDMSLRKTE---IDYEDIKTESIRLQOEYNNMLLTN 672

RESULT 6  
A33977  
myosin heavy chain, nonmuscle - chicken  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Gallus gallus (chicken)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002  
C:Accession: A33977; S06116; A43422  
R:Shohet, R.V.; Contt, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989  
A:Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my  
A:Reference number: A33977; MUID:90046668; PMID:2813355  
A:Accession: A33977  
A:Molecule type: mRNA  
A:Residues: 1-1959 <SHO>  
A:Cross-references: GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383  
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.  
Eur. J. Biochem. 184, 611-616, 1989  
A:Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in  
s.  
A:Reference number: S06116; NUID:90032648; PMID:2806244  
A:Accession: S06116  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 716-1008 <KAT>  
A:Cross-references: GB:X17589  
A:Note: this translation is not annotated in GenBank entry GGMHCFMHA, release 114  
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.

J. Cell Biol. 118, 1085-1095, 1992  
 A:Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate  
 A:Reference number: A43422; MUID:92381096; PMID:1512291  
 A:Accession: A43422  
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1900-1959 <HOD>  
 A:Experimental source: brush border  
 A>Note: sequence extracted from NCBI backbone (NCBIP:111947)  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP: coiled coil; hydrolase; methylated amino acid; nucleotide  
 F:84-764/Domain: myosin motor domain homology <MMOT>  
 F:174-181/Region: nucleotide-binding motif A (P-loop)  
 F:552-565/Region: actin binding #status predicted  
 F:626-640/Region: actin binding #status predicted  
 F:837-1936/Domain: coiled coil #status predicted  
 F:837-1277/Region: S2  
 F:1278-1959/Region: light meromyosin  
 F:1937-1959/Domain: carboxyl-terminal <CBT>  
 F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
 F:180/Binding site: ATP (Lys) #status predicted  
 F:694,704/Active site: Cys #status predicted

Query Match 4.1%; Score 127.5; DB 1; Length 1959;  
 Best Local Similarity 19.8%; Pred. No. 6.1;  
 Matches 110; Conservative 79; Mismatches 177; Indels 189; Gaps 22;

QY 42 ELSKILAKLE-RLKQONEDLRMAESLRIPGPIQGTATGRVRLVLEQLVKAKEQIR-- 98  
 DB 1066 ELQQAIAELKILQSKKEELQ--ALARVEEAAQNNWALKIRLESQITELQEDLSE 1123  
 QY 99 ----NYKQARNDLGDHILRRRIEN-----GAKELWFFLQSELKLLKLEGNELQRH 148  
 DB 1124 RASRNKAERKQKDLGEELEALNTELEDTLDSAAQELRSKEQEVTVLKKTLEDEAKTH 1183  
 QY 149 ADEILLDLGHHSRIMTDLYL--SOTDAGAGWEKREKAKDLTELVRRIITVQLNPDKCSKAR 208  
 DB 1184 EAQI-----QEMRQKHSQAIEELAEQ-----LEQT 1209  
 QY 209 KLVCNKNGCGYCGQLHHVVCYFMAYGTORTLILLESQNRWATGWTVPVSETCTD 268  
 DB 1210 RVANLEKA-----KQALESER-----AELSNE 1232  
 QY 269 RSLGSLGHSGBVQKQNV--QVVELPIVDSLHPRPPYLPVLPEDLADRL--LRVHGDP 324  
 DB 1233 VKVLLQKGDAAEHRKQKVDQALQELQVKFTEGER-----VKTELAERVNKLQVELDNV 1285  
 QY 325 VNWVSQ----FVKYLIRPOPWLERI-----EETTKLGFKHPVIGVHVRTRDKVGT 372  
 DB 1286 TGLNQSDSKSIK-LAKDFALESALQDTQELLQBELRLKLSFS-----TKLKQT 1334  
 QY 373 EAAFPPIEYVMVHVEHFQLLERRMKVDRKRVYLATDDPSLLKKAETKYSNVEFISDMSI 432  
 DB 1335 EDEKALKEQLEEEBAKRNLEKQISVLQQAV-----EARKQWD----- 1374  
 QY 433 SWSAGLHNRYSNLSRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQTLHPDASANPHS 492  
 DB 1375 -----DLGEC-----LETAEEAKKKLQKD-----LES 1396  
 QY 493 LDDIYFGQNAHNIQVYVPHQPTKEIPMEPGDIIGVAGNHNGYSGKGNRKLKGTGL 552  
 DB 1397 LTQRY-----EETIAAYDKLEKTKRLRLQQLDLD--IAVDLDHQRTVSNLEKKQK--- 1445  
 QY 553 YPSYKVRKEITVKY 567  
 DB 1446 PDQLAAEKNISAKY 1460

## RESULT 7

S28061

SCP1 protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999

C:Accession: S28061  
 R:Meuwissen, R.L.J.; Offenberger, H.H.; Dietrich, A.J.J.; Riesewijk, A.; van Iersel, M.; H  
 ENBO J. 11, 5091-5100, 1992  
 A:Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase  
 A:Reference number: S28061; MUID:93099884; PMID:1464329  
 A:Accession: S28061  
 A:Molecule type: mRNA  
 A:Residues: 1-946 <MEU>  
 A:Cross-references: EMBL:X67805; NID:g57212; PIDN:CAA48006.1; PID:g57213  
 C:Genetics:  
 A:Gene: SCPI  
 C:Keywords: DNA binding

Query Match 4.1%; Score 127; DB 2; Length 946;  
 Best Local Similarity 20.3%; Pred. No. 2.5;  
 Matches 105; Conservative 86; Mismatches 188; Indels 138; Gaps 23;

QY 42 ELSKILAKLEKQONEDLRMAESLRIPGPIQGTATGRVRLVLEQLVKAKEQIENY 100  
 DB 104 ENKVSILKEEIQENKDL-----IKENNATRHWCNLLKELTCARSAEKTSKY 150  
 QY 101 K-----KQARNDLGD-----HEILRRRIENGAKELWFFLQSELKLLKLEGNELQR 147  
 DB 151 EYERETROVYDUNNIEKIMILAFEEELRVOAENARLEMHFKLKHEDHEKIQHLE-EEYQK 209  
 QY 148 HAD-----EILLDLGHHSRIMTDLYL--SOTDAGAGWEKREKAKDLTELVRRIITVQL 199  
 DB 210 EVNNKENQVSLLLIQSTEKENMKMDLTFLLSESRDKANQLEEK-----TKLQDENLKLKN 264  
 QY 200 NPKDCSKARKLVCNKNGCGYCGQLHHVVCYFMAYGTORTLILLESQNRWATGWTVPV 259  
 DB 265 EKQDLHLS-----ELEDIKMSMQRSMTQKTL---EEDLQIAT---KTIY 303  
 QY 260 RPVSE-----TCTDRSGLSTGHM----- 277  
 DB 304 QLTEKEAQAEEELNKAETHSLVTELKATCTCTLEELLRTTEQOENEDDLKITMELQ 363  
 QY 278 --SGEVKD---KNVQVVELPIVDSLHPRPPYLPVLPEDLADRLLRVHG-DPAVWVVSQ 330  
 DB 364 KKSSELEEMTKFNKKEVELEELKTI-----LAEQKLLDEKKQVEKLAEBELQKQEQ 415  
 QY 331 FVKYLI--RPOPWLERIEETTKLGFKHPVIGVHVRTRDKVGTAAAFHPPIEYVMVHVEE 388  
 DB 416 ELTELLQTRKEITHDLVEQVTVTKTSEH-----YLKQVEEMKTELEKEKLNIELTANS 470  
 QY 389 HFQLLERRMKVDRKRVYLATDDPSLLKKAETKYSNVEFISDMSISWSAGLHNRYSNLSR 448  
 DB 471 DMLLE-----NKKLVQESDMVLELKHQEDDIINCKQEBMLKQIETLEEK--EMNLR 523  
 QY 449 GVILDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480  
 DB 524 DELESVRKEFTQGDVEVKCKLDKSEENARSIEYEVLK 560

## RESULT 8

A61231

myosin heavy chain nonmuscle form A - human

N:Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A

N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Homo sapiens (man)

C:Date: 12-May-1994 #sequence\_revision 14-Jul-1994 #text\_change 19-Apr-2002

C:Accession: A61231; A34876; I52562; I61692

R:Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adeletstein,

Circ. Res. 69, 530-539, 1991

A:Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different

A:Reference number: A61231; MUID:91316803; PMID:1860190

A:Accession: A61231

A:Molecule type: mRNA

A:Residues: 1-715 &lt;SIM&gt;

A:Cross-references: GB:M69180; NID:g189029; PIDN:AAA61765.1; PID:g189030

R:Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.

Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990

A:Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through altern





**Query Match**      3.9%; Score 121.5; DB 2; Length 3187;  
**Best Local Similarity** 19.0%; Pred. No. 29;  
Matches 108; Conservative 76; Mismatches 177; Indels 207; Gaps 24;

QY 27 GHVLRNDHPDHSRELSKTLAKLERLKQONEDLRRMAESLRIPGFIQGATGRVRVL 86  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 1997 GNLAQAVEHHKKAAELSSFKILL-DTQSEARVLADNLK-----L 2037

QY 87 EEQLVKAKEIQENYYKQARNDLKGKHIELRRIENGAKELWFFLOSELKKKLKEGLEIQQ 146  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2038 KKELOSNGEKSIQSIIQO-----KDDELLRR-LEGAEEKHRK-EKNQM 2078

QY 147 RHAEITLLDGCHHERSTMTDYLYSQTGDAGWEKEAKDTLVQ-----R 193  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2079 EKLDAHREKAHVETDLAIEQVSIRTKD-----XMKELQSLSDTLQAALFTK 2128

QY 194 RITYLNPKDCSKARKLCVNINKGCYGCOLHHVVCFMIAYGTQRILTILESONRYATG 253  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2129 SMSLIQDDR-----RVIDEAKKWEOREG 2152

QY 254 G----NETFVPVSETCTDRSGSLTGHWGEVDKNNVVVELPIVDSLPHRPPLPVLP 309  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2153 DAIOKTCEEVRLKEENT-----ALKDQLROM-TIH-----M 2183

QY 310 EDLADRLLRHVGDPANWWWSOFKYILIRPOPWLIEREETTCKLGFKHPGVGHVRTDK 369  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2184 ELKIIVTSRLHEHDKEIW-----ESKAOTELQH-----QQKAYDK 2217

QY 370 VGYFAAFPLEEYVMVVEEHFOELLRMKVDDKRVRVLATDDPSLLKEAKTKSYNYBFISD 429  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2218 LQEEN-----KELMSOLEEAQGLYH-----DSK-----NELTKLESLSKLD 2255

QY 430 NTSWSAGLIN-RYTENS LGVILDTHFLSQADFVCTPSSFQVCV-----VAVIMOTLHP 484  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2256 QSTDLANSLKCEKREHNNEGIIKQ---OEADIQNCKFNCEQLETDLTASRELTTRLHD 2311

QY 485 DASANTFHSLDDIYFGGNAHNQIAVYP-HOPRTKEIMPPEGDIIGVAGNHWNGYSKGV 543  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2312 EINVKEQKI--ISLGSKEBAIQVAIALHQSHKEIKELE-----NLLSOEE 2357

QY 544 NRKL-----GKTGLPYSPYKVRKETVK 566  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 2358 EENLTLEEENKRAVEKTNQLTEAITIK 2385

RESULT 13  
A70387  
conserved hypothetical protein aa\_1006 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 02-Feb-2001  
C:Accession: A70387  
R.; Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; et al.  
Nature 392, 353-358, 1998  
A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; PMID:98196666; PMID:9537320  
A:Accession: A70387  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-978 <AQF>  
A:Cross-references: GB:AE000718; MID:g2983504; PIDN:AAC07092.1; PID:g2983515; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: aa\_1006  
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

**Query Match**      3.9%; Score 120.5; DB 2; Length 978;  
**Best Local Similarity** 18.2%; Pred. No. 6.9;  
Matches 94; Conservative 80; Mismatches 161; Indels 181; Gaps 18;

QY 34 DHPDHRSRELSKTLAKLERLKQONEDLRRMAESL-----RIPEG 72  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :

```
Db 500 RUT-----TORHEI-ESTN----- 512
Qy 273 STGWSGEVKDKNVQVLP-IVDSLHPRPPLPLAVPE--DLADRLRV-----HGDDPA 324
Db 513 -----KSELRIAEITHLQQQLQSSQQLGKMIPEKQSLIDQLQVQNSLHRDSL 563
Qy 325 VVWSQFVKYLRPQWLEREETTKLGFHPVIGVHR-RTDKV--GTEAAPHPIEE 381
Db 564 L-----TLKRALE--TKEIGRQ-----QLRQLDEVEKETRAKLQEIDV 600
Qy 382 YNVVHEHFQLLERRM-----KVDKKVYLATDDPSLLKEAKTKSYNYEISDNS 431
Db 601 FNNQLKELRELKYNKQFQKQDFETEKIKQKELERKTSSELDKLBKEDKR-----RMLQDOK 656
Qy 432 ISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFH 491
Db 657 L-W-----QDRVKQEEERYKFQDEEKEREE-----SVQKCEV--EKKPEIQEKNKPPH 703
Qy 492 SLDDIYFQGG-----NAH-----NO-----IAYVPHQPRTKKEEIPMBPGDIIGVAG 533
Db 704 QPPEFGKLGGOIPWMTTEKAPLTINQGDVKKVYRYALYPFDARSHDEITIEPBGDIIMVDE 763
Qy 534 NH-----WNGYSKGVNRKLGKGLYPSYKVKREKISTVKYP--TYPEAE 574
Db 764 SQTGFGMLG-----GELKGTGWFFA-NYAERMPSEFFPSTTKPAE 805

RESULT 15
B43402
myosin heavy chain-B, neuronal - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: B43402; A43402
R:Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A:Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A:Reference number: A43402; MUID:92388144; PMID:1355479
A:Accession: B43402
A:Molecule type: mRNA
A:Residues: 1-2007 <TAK>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48988.1; PID:G212452
A>Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A:Accession: A43402
A:Molecule type: mRNA
A:Residues: 1-211;222-631;653-2007 <TA2>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48985.1; PID:G212449
A>Note: sequence extracted from NCBI backbone (NCBIN:112864)
C:Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F:1-211;222-631;653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS
F:88-802/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:212-221/Region: alternatively spliced segment 1 #status experimental
F:559-593/Region: actin binding #status predicted
F:632-652/Region: alternatively spliced segment 2 #status experimental
F:692-714/Region: actin binding #status predicted
F:875-2007/Domain: coiled coil #status predicted <COI>
F:875-1315/Region: S2
F:1316-2007/Region: light meromyosin
F:129/Modified site: N6 N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:732,742/Active site: Cys #status predicted
F:1954/Binding site: phosphate (Thr) (covalent) #status predicted
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.9%; Score 119.5; DB 1; Length 2007;
Best Local Similarity 24.2%; Pred. No. 21;
Matches 44; Conservative 37; Mismatches 46; Indels 55; Gaps 8;
```

```
Qy 40 SRELSKILAKLERLKQONEDLRRMAESLRIPGPIDQGTATGRVVRVLEQLVKAKEQIEN 99
Db 1359 TRQKLNLSRRIRQLREBEKNLQEQOE-----EEEEAR---KNLEKQMLAQOQLAE 1406
Qy 100 YKQARNDLG-----KDHEILRRRIENGAKELWFFLQSELKKLCKLEGNELQ 146
Db 1407 AKKVDDDLGTTIEGLEENKKLLKDMESLSORLEKA-----MAYDKLEKTK---NRLQ 1457
Qy 147 RHAEIILLDLGHHSIMTDL-----YYLSQTDGA-GEWREKEAKD 186
Db 1458 QELDDLAVDL-DHQRIQVSNLEKKQKKFQDMLAEEKNISARYAEERDRAEAAREKETKA 1516
Qy 187 LT 188
Db 1517 LS 1518

Search completed: February 2, 2004, 08:43:55
Job time : 21 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 10.5 Seconds  
(without alignments)  
2575.272 Million cell updates/sec

Title: US-09-971-773-23  
Perfect score: 3081  
Sequence: 1 MRAWTGSWRIMLILFWANGT.....YKVRKIETVKYTPAEAK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3013	97.8	575	1	FUT8_MOUSE
2	2984	96.9	575	1	Q9bys2 m alpha-(1,
3	2939	95.4	575	1	Q9n0w2 b alpha-(1,
4	2885	93.6	575	1	P79282 s alpha-(1,
5	1333	4.3	1727	1	Q9utk5 schizosacch
6	130.5	4.2	751	1	Q9vj3 drosophila
7	127.5	4.1	1959	1	P14105 gallus gall
8	127	4.1	997	1	Q03410 rattus norv
9	124.5	4.0	845	1	Q60563 mesocricetu
10	124.5	4.0	1960	1	P35579 homo sapien
11	124.5	4.0	1976	1	Q27991 bos taurus
12	120.5	3.9	978	1	O67124 aquifex aeo
13	119.5	3.9	1270	1	O42287 xenopus lae
14	118.5	3.8	976	1	Q15431 homo sapien
15	118.5	3.8	1427	1	Q310622 homo sapien
16	118	3.8	1217	1	Q9wve9 rattus norv
17	116.5	3.8	1961	1	Q62812 rattus norv
18	116	3.8	214	1	Q92882 homo sapien
19	116	3.8	993	1	Q62209 mus musculu
20	116	3.8	2442	1	Q9bv73 homo sapien
21	115.5	3.7	3680	1	Q97592 canis famil
22	115	3.7	896	1	Q9hlr8 thermoplasm
23	115	3.7	1027	1	P33175 mus musculu
24	115	3.7	1976	1	P35580 homo sapien
25	115	3.7	3210	1	P49454 homo sapien
26	114.5	3.7	2017	1	Q99323 drosophila
27	114	3.7	1976	1	Q91lt0 rattus norv
28	112.5	3.7	1935	1	Q90339 cyprinus ca
29	112.5	3.7	1938	1	P13538 gallus gall
30	112	3.6	2415	1	P08032 mus musculu
31	111.5	3.6	880	1	Q9uzc8 pyrococcus
32	111.5	3.6	919	1	Q9vzf1 aeropyrum p
33	111.5	3.6	2230	1	Q11439 homo sapien

## RESULT 1

ID	FUT8_MOUSE	STANDARD;	PRT;	575 AA.
AC	Q9WTS2; Q921U1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Alpha-(1-6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-fucosyltransferase)			
DE	(GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)			
DE	(alpha1-6FucT) (Fucosyltransferase 8)			
GN	FUT8			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=20358720; PubMed=10902914;			
RA	Hayashi H., Yoneda A., Asada M., Ikekita M., Imamura T.;			
RT	"Molecular cloning of mouse alpha-1,6-fucosyltransferase and			
RT	expression of its mRNA in the developing cerebrum.";			
RL	DNA Seq. 11:91-96 (2000).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McWen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fabey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,			
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).			
CC	FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to			
CC	the first GlcNAc residue, next to the peptide chains in N-glycans			
CC	(By similarity).			
CC	!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-{N-acetyl-beta-D-			
CC	glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-			
CC	glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-			
CC	>4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-			
CC	glucosaminyl]asparagine = GDP + N(4)-{N-acetyl-beta-D-			
CC	glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-			

## ALIGNMENTS

CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
 CC N-acetyl-beta-D-glucosaminyl]asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
 CC form in trans cisternae of Golgi (By similarity).  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB025198; BAA76392.1; -;  
 CC DR EMBL; BC010666; AAH10666.1; -;  
 CC DR MGD; MGI:1858901; Fut8.  
 CC DR InterPro; IPR001452; SH3.  
 CC DR Pfam; PF00018; SH3; 1.  
 CC DR SMART; SM00326; SH3; 1.  
 CC DR PROSITE; PSS0002; SH3; FALSE NEG.  
 CC KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 CC Golgi stack; SH3 domain.  
 CC FT DOMAIN 1 9  
 CC FT TRANSMEM 10 30 CYTOPLASMIC (POTENTIAL).  
 CC FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 CC FT (POTENTIAL).  
 CC FT LUMENAL, CATALYTIC (POTENTIAL).  
 CC FT DOMAIN 31 575  
 CC FT SH3.  
 CC FT SITE 299 305 SH3-BINDING (POTENTIAL).  
 CC FT SITE 365 366 IMPORTANT FOR DONOR-SUBSTRATE BINDING.  
 CC FT CONFLICT 40 40 T -> S (IN REF. 2).  
 CC FT CONFLICT 388 388 Q -> E (IN REF. 2).  
 CC FT CONFLICT 418 418 N -> K (IN REF. 2).  
 CC FT CONFLICT 418 418 N -> K (IN REF. 2).  
 CC SQ SEQUENCE 575 AA; 66555 MW; 7BE2ED3146E0C45F CRC64;  
 CC -----  
 CC Query Match  
 CC Best Local Similarity 97.8%; Score 3013; DB 1; Length 575;  
 CC Matches 558; Conservative 11; Mismatches 6; Indels 0; Gaps 0;  
 CC -----  
 CC QY 1 MRAWTGSRWIMLILFAWGTLLFYGGHLVRNDHPDHSRSLSKLAKLERLQONEDL 60  
 CC DB 1 MRAWTGSRWIMLILFAWGTLLFYGGHLVRNDHPDHSRSLSKLAKLERLQONEDL 60  
 CC QY 61 RMAESLRIPGPIDQGTATGRVLEEQLVKAKEQIENYKQARNGLGKQHEILRRRIE 120  
 CC DB 61 RMAESLRIPGPIDQGTATGRVLEEQLVKAKEQIENYKQARNGLGKQHEILRRRIE 120  
 CC QY 121 NGAKELWPFQSELKKLKLEGNELQRADEILDLGHHERSIMTDLVYLSOTDCAGEWR 180  
 CC DB 121 NGAKELWPFQSELKKLKLEGNELQRADEILDLGHHERSIMTDLVYLSOTDCAGEWR 180  
 CC QY 181 EKEAKDLTELVRRTYTLQNPDCSKARKLVNKNKGGYQGLHHVYCFMAYGTQRT 240  
 CC DB 181 EKEAKDLTELVRRTYTLQNPDCSKARKLVNKNKGGYQGLHHVYCFMAYGTQRT 240  
 CC QY 241 LILESQNRWYATGHWETVRPVSCTDRSGLSGTGHWSEGVKQVQVVELPVDLSLHPR 300  
 CC DB 241 LILESQNRWYATGHWETVRPVSCTDRSGLSGTGHWSEGVKQVQVVELPVDLSLHPR 300  
 CC QY 301 PPYLPLAVPEDLADLLRVRHGPVAVWVSQFVKYLRPOPMLEREIEETTKLGFKHPVI 360  
 CC DB 301 PPYLPLAVPEDLADLLRVRHGPVAVWVSQFVKYLRPOPMLEREIEATKLGKHPVI 360  
 CC QY 361 GVHVRRTDKVGTAAFPPIEEYVMVHVEHFQLLRMRKVKRKYLATDDPSLLKKAATK 420  
 CC DB 361 GVHVRRTDKVGTAAFPPIEEYVMVHVEHFQLLRMRKVKRKYLATDDPSLLKKAATK 420  
 CC QY 421 YSNYEFISDNTSWSAGLHNRYTENSRLGRVILDIHFLSQADFLVCTFSQVCRVAYEMQ 480  
 CC DB 421 YSNYEFISDNTSWSAGLHNRYTENSRLGRVILDIHFLSQADFLVCTFSQVCRVAYEMQ 480

QY 481 TLHPDASANFHSLLDDIYFQGNQAHNQIAVTPHQPTKEEIPMEPGDIIGVAGNHWGYS 540  
 DB 481 TLHPDASANFHSLLDDIYFQGNQAHNQIAVTPHQPTKEEIPMEPGDIIGVAGNHWGYS 540  
 QY 541 KGVNRKLGKTKGLYSYKVKREKIVKPYTPEAEK 575  
 DB 541 KGVNRKLGKTKGLYSYKVKREKIVKPYTPEAEK 575

## RESULT 2

FUT8\_HUMAN  
 ID FUT8\_HUMAN STANDARD; PRT; 575 AA.  
 AC Q9BYC5; O00235; Q9BYC6; Q9P2U5; Q9P2U6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
 DE fucosyltransferase) (GDP-fucose-glycoprotein fucosyltransferase)  
 DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
 DE (alpha1-6FucT) (Fucosyltransferase 8).  
 GN FUT8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.  
 EX MEDLINE=97279058; PubMed=9133635;  
 RA Taniguchi N.;  
 RA Yanagidani S., Uozumi N., Ihara Y., Miyoshi E., Yamaguchi N.,  
 RT "Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-  
 RT glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 FucT) from human  
 RT gastric cancer MKN45 cells.";  
 RL J. Biochem. 121:626-632(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Embryo;  
 RA Calileau A., Balanzino L., Candelier J.J., Oriol R., Mollicone R.;  
 RT "Differential splice variants of human FUT8 embryonic cDNA.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBAJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=20275614; PubMed=10814706;  
 RA Yamaguchi Y., Ikeda Y., Takahashi T., Ihara H., Tanaka T., Sasho C.,  
 RA Uozumi N., Yanagidani S., Inoue S., Fujii J., Taniguchi N.;  
 RT "Genomic structure and promoter analysis of the human  
 RT alpha1,6-fucosyltransferase gene (FUT8).";  
 RL Glycobiology 10:637-643(2000).  
 RN [4]  
 RP DONOR SUBSTRATE BINDING, AND MUTAGENESIS OF ARG-365 AND ARG-366.  
 RX MEDLINE=20229550; PubMed=10764839;  
 RA Takahashi T., Ikeda Y., Tateishi A., Yamaguchi Y., Ishikawa M.,  
 RA Taniguchi N.;  
 RT "A sequence motif involved in the donor substrate binding by  
 RT alpha1,6-fucosyltransferase: the role of the conserved arginine  
 RT residues.";  
 RL Glycobiology 10:503-510(2000).  
 CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
 CC the first GlcNAc residue, next to the peptide chains in N-glycans.  
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
 CC glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1-  
 CC >6)]-N-acetyl-beta-D-glucosaminyl]asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
 CC form in trans cisternae of Golgi (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:



CC in heart, lung, spleen and kidney.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF247186; AAF65460.1; -;  
 CC InterPro: IPR001452; SH3.  
 CC Pfam: PF00018; SH3; 1.  
 CC SMART: SM00326; SH3; 1.  
 CC PROSITE: PS00002; SH3; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; SH3 domain.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 30  
 FT  
 FT DOMAIN 31 575  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT SH3-BINDING (POTENTIAL).  
 FT SH3-BINDING (POTENTIAL).  
 FT SITE 299 305  
 FT SITE 365 366  
 FT IMPORTANT FOR DONOR SUBSTRATE BINDING (BY  
 FT SIMILARITY).  
 FT  
 FT SEQUENCE 575 AA; 66492 MW; FE04C66B8A5BF540 CRC64;  
 SQ  
 Query Match 95.4%; Score 2939; DB 1; Length 575;  
 Best Local Similarity 94.4%; Pred. No. 1.6e-188;  
 Matches 543; Conservative 15; Mismatches 17; Indels 0; Gaps 0;  
 QY 1 MRATGSGNRWIMLIFAWGTLFLFYGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
 DB 1 MRPTGSGNRWIMLIFAWGTLFLFYGGHLVRNDHPDHSRSLSKILAKLERLKQONEDL 60  
 QY 61 RMAESLRIPGEPIDQGTATGVRVRVLEQLVKAKEQIENYKQARNDLGKQHEILRRRIE 120  
 DB 61 RMAESLRIPGEPIDQGPASGRIRALEQLVKAKEQIENYKQTRNGLGKQHEILRRRIE 120  
 QY 121 NGAKELWFLQSELKLLKLEGNELORHADELILLGHHSRISIMTDLYLSQTDGAGWR 180  
 DB 121 NGAKELWFLQSELKLLKLEGNELORHADELFLSLGHHSRISIMTDLYLSQTDGAGWR 180  
 QY 181 EKEAKDLTVLORRITYLQNPXDCSKARKLVNCKGCGYQLHVVVCFMIAVGTQT 240  
 DB 181 EKEAKDLTVLORRITYLQNPXDCSKARKLVNCKGCGYQLHVVVCFMIAVGTQT 240  
 QY 241 LILESQNRVYATGGWETVPRVSETCTDRSGLSTGHSGEVKDKNVQVVELPVDLSLHPR 300  
 DB 241 LILESQNRVYATGGWETVPRVSETCTDRSGVYTGHSGEIKDKNVQVVELPVDLSLHPR 300  
 QY 301 PLYPLAVPEDLADLLRVHGDPAVWVSQFVKYLIRPQWLERIEETTKLGFKHPVI 360  
 DB 301 PLYPLAVPEDLADLLRVHGDPAVWVSQFVKYLIRPQWLERIEETTKLGFKHPVI 360  
 QY 361 GVHVRTDKVGTAAFPHTPEEYVHVEHFQLLERRMKVVKRVVLTATDDPSLLKEATK 420  
 DB 361 GVHVRTDKVGTAAFPHTPEEYVHVEHFQLLARRMQVVKRVVLTATDDPSLLKEATK 420  
 QY 421 YSNYEFISDINSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQ 480  
 DB 421 YPHYEFISDINSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFFSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYFPGQNAHQVAYPHQRTKEEIPMEPCDIIGVAGNHWGYS 540  
 DB 481 TLHPDASANFSLDDIYFPGQNAHQVAYPHQRTKEEIPMEPCDIIGVAGNHWGYS 540  
 QY 541 KGVNKLKGTGLYPSYKVKREKETVKYPTYPEAK 575  
 DB 541 KGVNKLKGTGLYPSYKVKREKETVKVHVPEAK 575

## RESULT 4

FUT8\_PIG  
 ID FUT8\_PIG STANDARD; PRT; 575 AA.  
 AC P79282;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
 DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)  
 DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
 DE (alpha1-6fuct) (Fucosyltransferase 8).  
 GN FUT8  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 102-130; 333-344; 402-415 AND  
 RP 566-575, AND FUNCTION.  
 RC TISSUE=Brain;  
 RX MEDLINE=97066976; PubMed=8910378;  
 RA Uozumi N., Yanagidani S., Miyoshi E., Ihara Y., Sakuma T., Gao C.-X.,  
 RA Teshima T., Fujii S., Shiba T., Taniguchi N.;  
 RT "Purification and cDNA cloning of porcine brain GDP-L-Fuc:N-acetyl-  
 RT beta-D-glucosaminide alpha1-->6fucosyltransferase.";  
 RL J. Biol. Chem. 271:27810-27817(1996).  
 CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
 CC the first GlcNAc residue, next to the peptide chains in N-glycans.  
 CC The pH optimum is 7.  
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-{N-acetyl-beta-D-  
 CC Glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
 CC Glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
 CC Glucosaminyl}asparagine = GDP + N(4)-{N-acetyl-beta-D-  
 CC Glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
 CC Glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
 CC N-acetyl-beta-D-glucosaminyl}asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
 CC form in trans cisternae of Golgi (By similarity).  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D86723; BAA13157.1; -;  
 CC InterPro: IPR001452; SH3.  
 CC Pfam: PF00018; SH3; 1.  
 CC SMART: SM00326; SH3; 1.  
 CC PROSITE: PS00002; SH3; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 KW Golgi stack; SH3 domain.  
 FT DOMAIN 1 9  
 FT TRANSMEM 10 30  
 FT  
 FT DOMAIN 31 575  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT SH3-BINDING (POTENTIAL).  
 FT SH3-BINDING (POTENTIAL).  
 FT SITE 299 305  
 FT SITE 365 366  
 FT IMPORTANT FOR DONOR SUBSTRATE BINDING (BY  
 FT SIMILARITY).  
 FT  
 FT SEQUENCE 575 AA; 66229 MW; 0F199D0BC2018F7B CRC64;  
 SQ  
 Query Match 93.6%; Score 2885; DB 1; Length 575;  
 Best Local Similarity 92.9%; Pred. No. 6.5e-185;

Matches 534; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 1 MRATGSRWMLILFAWGTLLFYIGGHLVRDNDHPDSSRELKILAKLRLKQONEDL 60  
 DB 1 MRPTGSRWMLILFAWGTLLFYIGGHLVRDNDHSSRELKILAKLRLKQONEDL 60  
 QY 61 RMAESLRIPGPDIDQGTATGRVRLBQLVKAKEQIENYKQARNDLGKHDEILRRRIE 120  
 DB 61 RMAESLRIPGPDIDQGTATGRVRLBQLVKAKEQIENYKQARNDLGKHDEILRRRIE 120  
 QY 121 NGAKELWFFLOSELKLLKLSNGLNQLRHADELILDLGHHSRIMTDLIYLSQTDGAGBWR 180  
 DB 121 NGAKELWFFLOSELKLLKLSNGLNQLRHADELILDLGHHSRIMTDLIYLSQTDGAGBWR 180  
 QY 181 EKEAKDLTELVRRTIYLNQPKDCSKARKVNCINKGCGYCOLHVVYCFMIAVGTORT 240  
 DB 181 EKEAKDLTELVRRTIYLNQPKDCSKARKVNCINKGCGYCOLHVVYCFMIAVGTORT 240  
 QY 241 LILESQNRWYATGWETVFRPVSCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300  
 DB 241 LALESHNRWYATGWETVFRPVSCTDRSGLSGTGHSGEVKDKNVQVVELPIVDSLHPR 300  
 QY 301 PPYLPLAVPEDLADRLVRHGDPAVWVVSQFVKYLIRPQPWLKEIEEATKGLGFKHPVI 360  
 DB 301 PPYLPLAVPEDLADRLVRHGDPAVWVVSQFVKYLIRPQPWLKEIEEATKGLGFKHPVI 360  
 QY 361 GVHVRTDKVTEAAPHPIEEMVHVEHFOLLERMKVDRKRVYLATDDPSLLKEATK 420  
 DB 361 GVHVRTDKVTEAAPHPIEEMVHVEHFOLLERMKVDRKRVYLATDDPSLLKEATK 420  
 QY 421 YSNTEFISDINSISWAGLNRYTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
 DB 421 YPSVEFISDINSISWAGLNRYTENSRLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
 QY 481 TLHPDASANFSLDDIYFGGQNAHQIAVPHQPTKEIPMEPGDIIGVAGNHNWYS 540  
 DB 481 ALHPDASANFSLDDIYFGGQNAHQIAVPHQPTKEIPMEPGDIIGVAGNHNWY 540  
 QY 541 KGVNRKLGKTGLPSYKREKTIETVKYTYPEAK 575  
 DB 541 KGVNRKLGKTGLPSYKREKTIETVKYTYPEADK 575

RESULT 5  
 ALMI\_SCHPO  
 ID ALMI\_SCHPO STANDARD; PRT; 1727 AA.  
 AC Q9UTK5; O13313; Q9UTR8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Abnormal long morphology protein 1 (Sp8).  
 GN ALMI OR SPAC1486.04C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,  
 Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
 Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 Goiffreau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 Daga R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;  
 "The genome sequence of *Schizosaccharomyces pombe*.";  
 Nature 415:871-880(2002).  
 [2]  
 SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=972;  
 RX MEDLINE=20123449; PubMed=10660053;  
 RA Jimenez M., Petit T., Gancedo C., Goday C.;  
 "The *alm1+* gene from *Schizosaccharomyces pombe* encodes a coiled-coil  
 protein that associates with the medial region during mitosis.";  
 Mol. Gen. Genet. 262:931-930(2000).  
 [3]  
 SEQUENCE OF 644-834 FROM N.A.  
 RC STRAIN=968 h90;  
 RX MEDLINE=20223868; PubMed=10759889;  
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,  
 Hiraoaka Y.;  
 "Large-scale screening of intracellular protein localization in living  
 fission yeast cells by the use of a GFP-fusion genomic DNA library.";  
 Genes Cells 5:169-190(2000).  
 RL Genes Cells 5:169-190(2000).  
 CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,  
 CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN  
 CYTOKINESIS.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL133357; CAB62414.1; -  
 DR EMBL; AF010473; AAB65416.1; ALT\_INIT.  
 DR EMBL; AB028012; BAA87316.1; -  
 DR PIR; T50073; T50073.  
 DR GeneDB SPombe; SPAC1486.04C; -  
 KW Coiled\_coil.  
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).  
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).  
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).  
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).  
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).  
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).  
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1727 AA; 137858 MW; F820BF8D9C132644 CRC64;  
 Query Match 4.3%; Score 133; DB 1; Length 1727;  
 Best Local Similarity 21.2%; Pred. No. 0.66;  
 Matches 110; Conservative 74; Mismatches 226; Indels 108; Gaps 20;  
 QY 34 DHPDSSRELKILAKLRLKQONEDLHMAESLRIPGPDIDQGTATGRVRLBQLVKA 93  
 DB 1166 NHVRKISEMEIISQROLLEFLFNKLRKRTVSSNRVLIADLQRG-----ITEKDVST 1218  
 QY 94 KEQI---ENY-----KKQARNDLGKHDEI---LRRRIENGAKELWPF----- 129  
 DB 1219 SESVGSRNLYNMVALLNESKNLSRENERNEEVITELREKTIETKTDLNAPRLNKEQLE 1278  
 QY 130 --LQSELKLLKLEGN--ELQRHADEILDLGHHSRIMTDLIYLSQTDGAGBWR 185

```

Db 1279 SOLQTEKAANKVLENSBEYKRHOEILLLSL-NSSTSTSSDASRLK-----NELVSKE-- 1330
Qy 186 DITELVQRRITVLPKPKDCSKARKLVINCNGKGYGCOLHHVVCYFMAYGTQRTLLILES 245
Db 1331 NLIELNQEIHLKSELETVKSKSDLENER-----AQNSKIEQLEL 1373
Qy 246 QNWRVATGQWETVPFVSETCTDRLSTGSHGSGVGVKQNVQVLPVDSLHPRPPVLP 305
Db 1374 KNTKLA-AAWRTKYQVNVKSKLEKNQIRQLQSRTSEAKVACHQNLQNLKPSATP 1432
Qy 306 LAVPEDLADRLRVHGDPAVWVSVFVKYLRPQPLWLEIEETTKLGFEXHVPVGHVR 365
Db 1433 TATTO-----SEPTSVSLEEFNS-----TKELSSTQKLSSEIMDLINTTKE 1474
Qy 366 RTDKV-----GTEAAAFH-PIEYIM-----VHVEHFQOLLERRMKVDKRVYLATDDPS 412
Db 1475 ELEKVRQNSKSEGTSKDTEIPNEEMERKVMQVQEVILRSRIAKELQKNELLRKQNOV 1534
Qy 413 LLKEAK-----TKYSNYEPIFSDNISWSAGLHNRVYNTENSLRGVLDIHF-----LSQA 460
Db 1535 LQDQVKALQETVWSSEAEASVHADTKOLENLKKTTEMLSVTFQVIFNESISDPSTSTA 1594
Qy 461 DELVCTFSSQVCRVAYEIMQTLHPDASANFH--SLDDI 496
Db 1595 DF--TTFVQKEWEKRREILQKDVEQVAQSHQKOLDNI 1630

RESULT 6
COG5_DROME
ID COG5_DROME STANDARD; PRT; 751 AA.
AC Q9GVD3; Q8T0G3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative conserved oligomeric Golgi complex subunit 5 (Four way stop
protein).
GN FWS OR CG6549.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RC STRAIN=Berkeley;
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Mozhrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M.; Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 406-751 FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacle J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -!- FUNCTION: Required for normal Golgi function (By similarity).
CC -!- SUBUNIT: Component of the conserved oligomeric Golgi complex which
CC is composed of eight different subunits and is required for normal
CC Golgi morphology and localization (By similarity).
CC -!- SUBCELLULAR LOCATION: Golgi (By similarity).
CC -!- SIMILARITY: BELONGS TO THE COG5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AE003655; AA053619.1;
CC EMBL; AY069335; AAL39480.1; ALT_INIT.
CC Flybase; FBGN0024689; fws.
CC Hypothetical protein; Transport; Protein transport; Golgi stack;
CC Membrane.
CC SEQUENCE 751 AA; 84953 MW; E32B1D05BEEF6EBA CRC64;
CC
CC Query Match 4.2%; Score 130.5; DB 1; Length 751;
CC Best Local Similarity 20.7%; Pred. No. 0.33;
CC Matches 127; Conservative 98; Mismatches 212; Indels 177; Gaps 32;
CC
CC Qy 32 DNDHPDSS-----RELKILAKLERLQKQNEEDLRRAESLRIPGP-IDOQTATGR 82
CC Db 19 DNDFTASMSHTIGQIQELSK-----QLQNTKEELH---QQVYDKHGALLQOATHAGR 69
CC Qy 83 ---VRVLEEOLVKAKE-----QIENYKQQRND---LQKHIEILRRRIENGAKELWF 128
CC Db 70 FDAALNALAEDVQVRVRETHGRLKNQVDTQYQVQVENQTVLGRHLDVSHLLRSAGT---- 124
CC Qy 129 FLOSELKKLKGLENEIQRHADEILLGLGH-HERSIMTDLYLSQTDGAGWEKEAKDL 187
CC Db 125 -LLSITAKLKATK--DVLRLA-EIHFELGQLIEDKELKDIDFIQOERAYVSSAKIKRL 180
CC Qy 188 TELVQRRTIYLO--NPKDCSKARKLVNCIN-----KGCYGCQL 224
CC Db 181 TQM--QLVTGQERNEQVNVNALKIFMNFNTLEKSLDNLNLLATFTADMEQSLKECFAGNDI 238
CC Qy 225 H-----HVYVCFMIAYGTQRT-LILESQNRVATGQWET-----VFRPVSETCTDRSGL 272
CC Db 239 SVLNKSPHNVSKPAPSRGPKGTQLTITQNR--AKFWKSLHWLLYLDLPETCTQIKLL 296
CC Qy 273 STG-----HWSGEVQKQNV-----QVVELPVDLSLHPRPYPPLAVPEDLADRL-- 317
CC Db 297 KTALEQINQFGYTSSESDQICPQRFQVQVQQLLRKSFDECPQHVTTQTLQEGLSKULTSAR 356
CC Qy 318 ---RVHGDPAVWVSVFVKYLRPQPLWLEIEETTKLGFEXHVPVGHVR-- 370
CC Db 357 GLEORLHGE-----FQDNELFAP-----LEVGVSVKCAANFKACLAGVDLPNGNETVDN 405

```

```
QY 371 -----GTEAPHPTEYVMHV-----BEHFOLLERRMKV--DKKRVVLATDDPSLL 414
DB 406 FIRVASTELSAALDSRLTNAIANVAFACGKELCTKLEAQIKLGADSKOV---VDLPNLQ 462
QY 415 KEATKYSNYBFISDINSWSAGLHNRVTENSLRGVLDIHF-----LS 458
DB 463 QQONTQLANVLF-----YKDSVRRMLSDLHVHPEKTPGTAREIISRSLE 507
QY 459 QADELVCTFFSQVCEVAYE---IMQTLHPDASANFHSLD-----DIYFGGQNAH 505
DB 508 QADLLIGHILOQIMESIITTSIIVLSMHREPLNSRMSTGTPSMYMKELQEFVNRWS 567
QY 506 NOIAYVPHQPTKE 519
DB 568 HHIALFDCKQMTKK 581

RESULT 7
MYH9_CHICK
ID MYH9_CHICK STANDARD; PRT; 1959 AA.
AC P14105;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle (Cellular myosin heavy chain) (NMWMC).
GN MYH9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=intestinal epithelium;
RX MEDLINE=90046668; PubMed=2813355;
RA Shohet R.V., Conti M.A., Kawamoto S., Preston Y.A., Brill D.A.,
RA Adelstein R.S.
RT "Cloning of the cDNA encoding the myosin heavy chain of a vertebrate
cellular myosin."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7726-7730(1989).
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CAPPING.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
REGULATORY LIGHT CHAIN SUBUNITS (ILC-2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26510; AAA48974.1; -.
DR PIR; A33977; A33977.
DR HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; Myosin head.
DR InterPro; IPR004009; Myosin N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; Myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
```

```
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSc; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 COILED COIL (POTENTIAL).
FT NP_BIND 837 1925 ATP (POTENTIAL).
FT NP_BIND 174 181 ACTIN-BINDING.
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1959 AA; 226502 MW; A75C86086FD3A1A1 CRC64;

Query Match 4.1%; Score 127.5; DB 1; Length 1959;
Best Local Similarity 19.8%; Pred. No. 1.8; Indels 189; Gaps 22;
Matches 110; Conservative

QY 42 ELSKILAKLE-RLKQONEDLRRMAESLRIPGPIDQGTATGRVRLVEELVKAKEQIE-- 98
DB 1066 ELQAIKLELQTSKEEELQ--AALARVEEAAQKMWALKKIRELSQITELQEDLESE 1123
QY 99 ----NYKQARNLDGKDHIEILRRRIEN-----GAKELWFFLQSELKKLKGELNQLRH 148
DB 1124 RASRNKAQKQKRDGLGELEALKTELEDTLDSTAQOELRSKREQEVTVLKKLEDEAKTH 1183
QY 149 ADEILLDLGHHERSIMTDLVYLSQTDGAGWEKEAKDLTELVCORRITYLQNPKDCSKAR 208
DB 1184 EAQI-----QEMRQKHSQAIELEAQ-----LEQTK 1209
QY 209 KLVCNINKGCGYQQLHHVVVCFMAYGTQRTLLLESQNWRYATGCGWETVPRVSETCTD 268
DB 1210 RVKANLEKA-----KQALESER-----AELSNE 1232
QY 269 RSGLTGHWGSEVXDKNV--QVVELPVDLSLHPPVPLVPLAVPEDLADRL--LRVHGDDA 324
DB 1233 VKVLQKGDGAEHKKKVDQAQLOELQVKTGEGER-----VKTLEAERVNKLQVELDNV 1285
QY 325 VWWYSQ----FVKVLIROPQWLEREI-----EETTKLGFKHPVGVHVRDKVGT 372
DB 1286 TGLNQSDSKSIK-LAKDFALESQLOTOELLQEETRLKLSFS-----TKLKQT 1334
QY 373 EAAPHPTEYVMHVVEHFQLLERRMKVDKKRVYLATDDPSLLKEAKTKYSNYEFTSDNSI 432
DB 1335 EDEKNALKEQLESEEAARNLEKQISVLQQAQV-----EARKQMD-----1374
QY 433 SWSAGLHNRVTENSLRGVLDIHFLSQADFLVCTFFSSQVCRVAYEIMQTLHPDASANFHS 492
DB 1375 -----DGLGC-----LEIAEEAKKKLQKD-----LES 1396
QY 493 LDDIYFGGQNAHNOIAVYPHQPTKEIPMEPGDIIIGVAGNHNGYSKGVNRKLGKTGL 552
DB 1397 LTQRY-----ESKIAAYDKLEKTKRLOQELDD-IAVDLDHORQTVSNLEKKQKK--- 1445
QY 553 YPSYKVRKEIETVXY 567
DB 1446 PDQLAEKKNISAXY 1460

RESULT 8
SCPI_RAT
ID SCPI_RAT STANDARD; PRT; 997 AA.
AC Q03410;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SYCP1 OR SCPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
```



```

RN  SEQUENCE FROM N.A.
RP  TISSUE=Testis;
RX  MEDLINE=93099884; PubMed=1464329;
RA  Meuwissen R.L.J., Offenber H.H., Dietrich A.J., Rieswijk A.,
RA  Iersel M., Heyting C.;
RT  "A coiled-coil related protein specific for synapsed regions of
RT  meiotic prophase chromosomes.";
RL  EMBO J. 11:5091-5100(1992).
CC  -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC  SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC  CHROMOSOMES DURING MEIOTIC PROPHASE.
CC  -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC  SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC  FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC  IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC  C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC  SYNAPTONEMAL COMPLEX.
CC  -!- TISSUE SPECIFICITY: Testis.
CC  -!- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE
CC  CELLS.
CC  -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC  FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC  DOMAIN HAS DNA-BINDING CAPACITY.
CC  -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC  CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE
CC  OTHER SPECIES SYCP1 SEQUENCES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X67805; CAA48006.1; ALT FRAME.
KW  Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW  DNA-binding; Coiled coil.
FT  DOMAIN 108 819 COILED COIL (POTENTIAL).
FT  DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  DOMAIN 982 990 ARG/LYS-RICH (BASIC).
FT  SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;
SQ
Query Match 4.1%; Score 127; DB 1; Length 997;
Best Local Similarity 20.3%; Pred. No. 0.8;
Matches 105; Conservative 86; Mismatches 188; Indels 138; Gaps 23;
QY 42 ELSKILAKLERLQKQNEEDLRMAESLRPEGPDQATGR-VRVLEEQLVKAKQIENY 100
DB 155 ENEKSVLKLEEEIQENKDL-----IKENNRTHWCNLLKTCARSABKTSKY 201
QY 101 K-----KQARNDLGRD-----HEILRRRIENGAKELWFFLOSELKXKLEGNELQR 147
DB 202 EVERETRQVYVDNNNIEKMILAFELRVAQENARLEHMFKLKEDHEKIQHLE-EEYQK 260
QY 148 HAD-----EILLDIGHERSIMTDLYL--SQTGAGEWRKEAKDLTELQVRITVLIQ 199
DB 261 EYNNKENQVSLLLIOSTEKENKMKDLTFLLSESRDKANQLEK-----TKLDQENLKEIN 315
QY 200 NPKDCSKARKLVNCKNGKGGYCOLHHVVCFMIAVGTORTLILESQNRWYATGGWTFV 259
DB 316 EKXDHLS-----ELEDIKMSQRSWSTOKTL--EEDLIQIAT-----KTIY 354
QY 260 RPVSE-----TCTDRSGLSTGHW----- 277
DB 355 QLTEEKEAQMEELNKAKTTHSLVVTLEKATCTCTLEELLRTQORLENNEDQLKLITMELQ 414
QY 278 --SGEVKD-----KNQVQVELPIVDSLHPRPPYLPVAVEDLADRLRLRVHG-DPAVWVWSQ 330
DB 415 KKSSELEEMTKFKNNEVELEBKTI-----LAEDQKLLDEKKQVEKLAEBELQKREQ 466

```

```

QY 331 FVKYLI--RPOFWLEREIEETTKLGFHPGVGHVRRTKVGTAAAFHPIEYVWVHVEE 388
DB 467 ELTFLQLQREKIHDLLEVQVTVTKTSEH-----YLKQVEEMKTELEKEKLNIELTANS 521
QY 389 HFOLLERRMKVDKRVYLATDPSLLKEAKTKYSNVEFISDINSISWAGLHNRVYTSNLSR 448
DB 522 DMLLE-----NKKLVQASDMVLKELKHQBDINDCKKOEERMLKQIETLEEK--EMNLR 574
QY 449 GVILDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480
DB 575 DELSVRKEFTQQGDEVKCKLDKSEENARSIEYEVLK 611

RESULT 9
SCPI_MESAU STANDARD; PRT; 845 AA.
ID SCPI_MESAU
AC Q60563;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome
DE synapctic protein) (Fragment).
DE SCPI OR SIN1.
OS Mesocricetus auratus (Golden hamster).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RX MEDLINE=95181577; PubMed=7876343;
RA Dobson M.J., Pearlman R.E., Karaiskakis A., Spyropoulos B.,
RA Moens P.B.;
RT "Synaptonemal complex proteins: occurrence, epitope mapping and
RT chromosome disjunction.";
RL J. Cell Sci. 107:2749-2760(1994).
CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE (BY SIMILARITY). HAS NON-
CC SPECIFIC DNA BINDING CAPABILITY.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX.
CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L32978; AAC42039.1; -.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT NON TER 1 672 COILED COIL (POTENTIAL).
FT DOMAIN <1 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 753 756 ARG/LYS-RICH (BASIC).
FT DOMAIN 830 838 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;
Query Match 4.0%; Score 124.5; DB 1; Length 845;
Best Local Similarity 18.8%; Pred. No. 0.95;

```



```

RESULT 10
MYH9_HUMAN
ID MYH9_HUMAN STANDARD; PRT; 1960 AA.
AC P35579; O60805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE MYosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiwich R., Beare D.M.,
RA Clamp M., Smink L.R., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagnuley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Evans P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Dhams K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grahm D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,

```

	EMBL; Z82215; CAB05105.1; -				
DR	EMBL; M81105; AAA59888.1; -				
DR	EMBL; M69180; AAA61765.1; -				
DR	EMBL; M11013; AAA36349.1; -				
DR	PIR; A61231; A61231.				
DR	HSP; P10587; IBR2.				
DR	Gnew; HGNC:7579; MYH9.				
DR	MIM; 160775; -				
DR	MIM; 153640; -				
DR	MIM; 155100; -				
DR	MIM; 603622; -				
DR	MIM; 605249; -				
DR	GO; GO:0005860; C:non-muscle myosin; TAS.				
DR	InterPro; IPRO00048; IQ_region.				
DR	InterPro; IPRO01609; myosin_head.				
DR	InterPro; IPRO04009; Myosin_N.				
DR	InterPro; IPRO02928; Myosin_tail.				
DR	InterPro; IPRO02017; Spectrin.				
DR	Pfam; PF00612; IQ; 1.				
DR	Pfam; PF00083; myosin_head; 1.				
DR	Pfam; PF02736; Myosin_N; 1.				
DR	Pfam; PF01576; Myosin_tail; 1.				
DR	PRINTS; PR00193; MYOSINHEAVY.				
DR	ProDom; PD000355; myosin_head; 1.				
DR	SMART; SM00015; IQ; 1.				
DR	SMART; SM00242; MYSC; 1.				
DR	PROSITE; PSS0096; IQ; 1.				
DR	Myosin; ATP-binding; Calmodulin-binding; Actin-binding;				
KW	Coiled coil; Alkylation; Multigene family; Disease mutation;				
KW	Deafness..				
FT	DOMAIN 1	778			
FT	FT DOMAIN	779	808		
FT	DOMAIN	837	1926		
FT	NP_BIND	174	181		
FT	DOMAIN	654	676		
FT	MOD_RES	694	694		
FT	VARIANT	704	704		
FT	VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT	1	778		
FT	FT FT	779	808		
FT	FT FT	837	1926		
FT	FT NP_BIND	174	181		
FT	FT MOD_RES	694	694		
FT	FT VARIANT	704	704		
FT	FT VARIENT	93	93		
FT	FT FT				

FT	VARIANT	702	702	/FTid=VAR 010791. R -> C (IN FINS).
FT	VARIANT	705	705	/FTid=VAR 010792. R -> H (IN DFNA17).
FT	VARIANT	1155	1155	/FTid=VAR 010793. T -> I (IN MHA).
FT	VARIANT	1165	1165	/FTid=VAR 010794. R -> C (IN SBS).
FT	VARIANT	1424	1424	/FTid=VAR 010795. D -> H (IN FINS).
FT	VARIANT	1841	1841	/FTid=VAR 010796. E -> K (IN MHA).
FT	CONFLICT	53	55	/FTid=VAR 010797. EAI -> RGH (IN REF. 3).
FT	CONFLICT	660	660	T -> S (IN REF. 3).
FT	CONFLICT	869	869	T -> M (IN REF. 4).
FT	CONFLICT	931	931	C -> Y (IN REF. 4).
FT	CONFLICT	1240	1241	KG -> GR (IN REF. 4).

Query Match 4.0%; Score 124.5; DB 1; Length 1960;  
Best Local Similarity 19.0%; Pred. No. 2.8;  
Matches 105; Conservative 95; Mismatches 191; Indels 161; Gaps 26;

Qy	32	DNDPHPSRELSKILAKLERLKQ	-----NEDLRMAESLRIPGEGIDOGTATGRVRVLE	87
Db	1054	EGDSTLSD-QIAELQAIKLMQLAKKEEELQ	--AALARVEEAAQNMALKKIRELE	1110
Qy	88	EQLVKAKEQIE	-----NYKKQARNLDGKHDIRRRRIENGAKELWFFQLSESLKKLKLLE	141
Db	1111	SQISELQEDLESERASRNAEKQKRDGLGELKALTELEDTLDST	---AAQELRSKREQE	1168
Qy	142	GNEQLRHADEIILLDGHGHERSIMTDLVYLSQTDGAGEMREKAKOLTELVRRIITYLQNP	201	
Db	1169	VNLLKKTLEE--EAKTHEAQIQ	-----EMEQKHSQAVEELAEQ	1204
Qy	202	KQCSKARKVLVNCINKGCGYGCOLHHVVCYCFMAYGTQRTLLILESQNWRYATGGWETVFRP	261	
Db	1205	--LEQTKRVKANLEK	-----AKQTL-----ENER	1226
Qy	262	VSETCTDRSGLSGHWSGVEKQNV--QVVELPIVDLSLHPRPPYLPPLAYPEDLADRLLRV	319	
Db	1227	-GELANEVKVLLQKGDSEHKRKKVBAQQLQVKFNEGER	-----VRTLEADKVTK-	1277
Qy	320	HGDPAAWVWSQFVKYLIRPQPWLREIEETTKLGFHPVIGVHVHVRTDKVGTEAA--FH	377	
Db	1278	-----LQVELDNTGLS	-----QSDSKSKLTKDFS	1304
Qy	378	PIEYVHVVEEHFOLLERMKVDDKRVYATDDPSLLKKAATKYSNYEFISDINSISWSAG	437	
Db	1305	ALBSQLQDTQELLQ	-----EENRQKLSLSTK-----LKQ-----VEDEKNSFREQ	1344
Qy	438	L-HNRYTENSIRGVILDIHFLSQADFLVCTFSQVC-RVAYEIMOTLHPDASANFSLDD	495	
Db	1345	LEEBEBAKNLEKQIATLH-AQVADMKKKQWEDSVGCLETAEEVKRLQKQD	-----LEG	1396
Qy	496	IYVFGQNAHQIAVYPHQPRTKEEIPMEPGDIIIGVAGNHNWNGSYGKGNRKLKGTGLYPS	555	
Db	1397	L-----SQRHEEKVAAVDKLEKTKYRLQOELDDL--VDLDHQRSACNLEKKQCK	---FDQ	1448
Qy	556	YKVRKXIETVKY	567	
Db	1449	LLAAEKTISAKY	1460	

RESULT 11  
MYHA BOVIN  
ID MYHA BOVIN STANDARD; PRT; 1976 AA.  
AC Q27951;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,  
DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).



```
QY 73 PIDQGTATGRVRLBEEVLKAKEQI-----ENYKQARNDLGKDH----- 112
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 KVRKNKLTRELAVLKDELSTFAQELNRIEAEKEKEKEKELEHRLKLLKQLKEITL 353
QY 113 -----EILRRINGAKELMFFLOSELKLLKLEGN----- 143
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 ELSQLSSSLKEREYEQAQFEDLSRVERKG-KKLVAETEEKLEKELFSEBEYTS 412
QY 144 -----ELQRHDEILLDLGHHSIMTDLYLVSQTDGAGWEKEAKDLTELVRRI 195
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 KKKERLLVELQRLKELKEKEQLEN--LTQKI-----KEKKVHEKVLNEL 457
QY 196 TYLQNPDKCSKARKLVNCKNGCGYCOLH-HVVYCFM-----IAYGTQRTLILESON 247
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 KELERE--LKEREL-----HYHAHWASVLSPGDTCPCVCGGIYRGKALEN-- 500
QY 248 WRYATGGWETVRPVSCTDRSGSLSTGHWSGEVDKQNVQVVELPI-----V 294
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 -----VDAEGISELKHAKELKEKEBEREIDTTLKLYAQKINSLEKEEM 541
QY 295 DSIHPRPPVPLAVPEDLADRL-----LRVHGDPVAVWVSQFVKYLIRPOPWLER----- 344
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 EKLRNEVEELRKEIPENLKERIKLLEKRIEKELEHKLKLYRKALEDROKQKEAQA 601
QY 345 -----EIEETTKKLGKPHVIG-----VHVRRTDKVG-----TEAAPHPIEYVMVH 385
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 602 HKAQTELELLKELKIREKSLVKFEKLYRVERLEDYESSLKEEINVINSKLEIEE 661
QY 386 VEEHP-OLLERMVNDKRVLYATDDPSLLKKAATK 420
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 LRKHFEELSSRSKLEGEALSALNESINSLEERKEK 697
```

```
RESULT 13
ITNI_XENLA
ID ITNI_XENLA STANDARD; PRT; 1270 AA.
AC O42287;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Intersectin 1.
GN ITSN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=99030416; PubMed=9813051;
RA Yamabhai M., Hoffman N.G., Hardison N.L., McPherson P.S.,
RT Castagnoli L., Cesareni G., Kay B.K.;
RT "Intersectin, a novel adaptor protein with two eps15 homology and five
RT src homology 3 domains.";
RL J. Biol. Chem. 273:31401-31407(1998).
CC -!- FUNCTION: Adaptor protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC Enriched in synaptosomes (By similarity).
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- SIMILARITY: Contains 5 SH3 domains.
CC -!- SIMILARITY: Contains 2 EH domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AF032118; AAC73068.1; -
DR PIR; T09194; T09194.
DR HSSP; P29355; 1SEM.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15 homology.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehand; 2.
DR PRINTS; PR00018; SH3; 5.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 5.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00018; EF HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50002; SH3; 5.
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding.
KW DOMAIN 21 109
FT CA BIND 66 77 EF-HAND 1 (POTENTIAL).
FT DOMAIN 220 309 EH 2.
FT CA BIND 266 277 EF-HAND 2 (POTENTIAL).
FT DOMAIN 325 697 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
FT DOMAIN 732 793 COILED COIL (POTENTIAL).
FT DOMAIN 897 955 SH3 1.
FT DOMAIN 986 1044 SH3 2.
FT DOMAIN 1058 1122 SH3 3.
FT DOMAIN 1139 1198 SH3 4.
FT DOMAIN 1270 AA; 143670 MW; EA940C1F6B6A6858 CRC64;
SQ SEQUENCE 1270 AA; 143670 MW; EA940C1F6B6A6858 CRC64;

Query Match 3.9%; Score 119.5; DB 1; Length 1270;
Best Local Similarity 20.6%; Pred. No. 3.5;
Matches 121; Conservative 82; Mismatches 174; Indels 211; Gaps 33;

QY 41 RELSKILAKLERLKQONEDLRMAESLRIPGPIDQGTATGRVRLBEEVLKAKEQIENY 100
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 RKEQERLAQLERAERKERERQDERK-----RQDLEKQLEKQRE-LERQ 420
QY 101 KQARNDLGKHDEILRRRIENGAKELMFFLOSELKLLKLEGNELQRHDEILLDLGH 160
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 REERR-----KEIERR-----EAEKR-----ELERQLEWERNR--QELL 456
QY 161 RSIWTDLYLSQTDGAGWEKEAKDLTELVRRI-----LQNPDKC--SKARKLVC 212
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 -----NQRNEQEDIVVLKAKKTLFEFEALNDKHKQLEGLQDIRC 499
QY 213 NINRGCGYCOLHHVYCFMIAYGRTLTILESQNVRYATCGWETVFRPVSETCTDRSGL 272
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 500 RL-----TORHEI-ESTN----- 512
QY 273 STGHWSGEVDKQNVQVVELPI-IVDSLHPRPPVPLAVPE--DLADRLRV-----HG 324
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 -----KSRELRIAEITHLQOQLQESQQLLGKMIPEKOSLIDOLKQVQONSILHRDSL 563
QY 325 VVWVSQFVKYLIRQPWLERIEETTKLKGFKHPVIGVHVR-RTDKV--GTEAAPHPIEE 381
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 564 L-----TLKRALE--TKEIRGQ-----QLRDQDLEVEKETRAKLEIDV 600
QY 382 YMVHVEEHFQLERRM-----KVDKRVVYLATDDPSLLKKAATKYSNYEYISNS 431
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 FNNQLKELRYNQKQFQKQDFETEKIKQELERKTSSELDKKEEDKR-----RMLEQDK 656
QY 432 ISWAGLHNRYTNSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQTLHPDASANFH 491
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 L-W-----QDRVKQBEERYKFDQBEKEKEE-----SVQKCEV--EKKEPQEKPNKPFH 703
QY 492 SLDDIYVFGGQ-----NAH-----NQ-----IAVYPHQPRTKKEIPMPGDIIGVAG 533
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 704 QPPRGLKGGQIPWNTKEKAPLTINQGDVKKVYVRYALYPPDARSHDEITIEFGDIIMWDE 763
```

QY 534 NH-----WNGYSGVNRKLGKGTGLYPSKYVREKIETVKYP--TYPEAE 574  
 Db 764 SQTGPGWLG-----GELKGTGWFFA-NYAERMPSEFPSTTKPAE 805

## RESULT 14

SCPI\_HUMAN  
 ID SCPI\_HUMAN STANDARD; PRT; 976 AA.  
 AC Q15431; O14963;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SYCP1 OR SCPI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97224467; PubMed=9119375;  
 RA Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J.,  
 RA Heyting C.;  
 RT "Human synaptonemal complex protein 1 (SCP1): isolation and  
 RT characterization of the cDNA and chromosomal localization of the  
 RT gene";  
 RL Genomics 39:377-384 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=98037449; PubMed=9371398;  
 RA Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,  
 RA Inazawa J., Takeo M., Nozaki M., Nojima H., Matsumiya K., Namiki M.,  
 RA Okuyama A., Nishimune Y.;  
 RT "Assignment of synaptonemal complex protein 1 (SCPI) to human  
 RT chromosome 1p13 by fluorescence in situ hybridization and its  
 RT expression in the testis";  
 RL Cytogenet. Cell Genet. 78:103-104 (1997).  
 CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: Testis.  
 CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X95654; CAA64956.1; --  
 CC EMBL; D67035; BAA22586.1; --  
 CC Genbank; HGNC:11487; SYCP1.  
 CC MIM; 602162; --  
 CC GO; GO:0005634; C:nucleus; TAS.  
 CC GO; GO:0005716; C:synaptonemal complex; TAS.  
 CC GO; GO:0003677; F:DNA binding activity; TAS.  
 CC GO; GO:0007131; P:meiotic recombination; TAS.  
 CC GO; GO:0007283; P:spermatogenesis; TAS.  
 CC GO; GO:0007129; P:synapsis; TAS.  
 CC Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 CC DNA-binding; Coiled coil.

FT DOMAIN 12 100 ASP/GLU-RICH (ACIDIC).  
 FT 107 798 COILED COIL (POTENTIAL).  
 FT 117 120 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT 679 682 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT 961 969 ARG/LYS-RICH (BASIC).  
 FT 46 46 L -> F (IN REF. 2).  
 FT 106 106 F -> Y (IN REF. 2).  
 FT 153 153 F -> C (IN REF. 2).  
 FT 161 161 K -> T (IN REF. 2).  
 FT 168 168 E -> D (IN REF. 2).  
 FT 216 216 N -> S (IN REF. 2).  
 FT 226 226 HG -> FE (IN REF. 2).  
 FT 350 350 K -> N (IN REF. 2).  
 FT 360 360 E -> D (IN REF. 2).  
 FT 400 401 KN -> NY (IN REF. 2).  
 FT 406 406 K -> I (IN REF. 2).  
 FT 415 415 K -> T (IN REF. 2).  
 FT 449 449 E -> D (IN REF. 2).  
 FT 483 510 IQLTAITTSQVYSKEVDLKTLENEK -> YSYCHYHKW  
 FT TVLPKRGQRPKLSSKRE (IN REF. 2).  
 FT LTSHCNKLSLENK -> YFTLQASPPN (IN REF.  
 FT 2).  
 FT 516 528  
 FT 549 549 N -> I (IN REF. 2).  
 FT 560 560 K -> T (IN REF. 2).  
 FT 805 805 E -> D (IN REF. 2).  
 FT 941 941 P -> S (IN REF. 2).  
 SQ SEQUENCE 976 AA; 111069 MW; 8BA81D042AC2696B CRC64;  
 Query Match 3.8%; Score 118.5; DB 1; Length 976;  
 Best Local Similarity 20.5%; Pred. No. 2.9;  
 Matches 90; Conservative 74; Mismatches 164; Indels 111; Gaps 19;  
 QY 42 ELSKILAKLEBLKQONEDLRMAESLRIPEDIGQTATGR---VRVLEEQLVKAKEQIE 98  
 Db 435 ELKVLGKETLLYENKQFEXIAELKGTQEQL-IGLLQAREKEVHDLLEIQLTATTSEQ 493  
 QY 99 NYKQARNDLCKDHEILRRRIENGAKELWFFLOSELKCLKLEGNELQORHADAELLDLGH 158  
 Db 494 YSKEVKD-----LATELEN-EKLNKLTSHCNKL-SLENKELTQETSDMTLELKN 543  
 QY 159 HERSI-----MTDLYLSQTD-----GAGEWREKEAKDLTELVRRTYLNQPKDC 204  
 Db 544 QQEDINNNKQEBRMKQIENLQETETQLRNELEYVREELKQKRDVCKLD--KSEENC 601  
 QY 205 SKARKLVGNIN-----KCGYGCQLHHVYCFMAYGQTORTLILLESQNRW 249  
 Db 602 NNLKQVENKNKYTEELQOENKALKKKGTAESKQLN--VYEIKV---NKLELELESQK 656  
 QY 250 YATGQWETVFRPVSETCTDRSGLSGTGHWGSKDKNQVQVVELPIVDLSLHPRPPYPLAVP 309  
 Db 657 F-----GEITDTYQKIEDKKIS-----E 675  
 QY 310 EDLADRL--LRVHSDPAV-----WVVSQVLYLIRPQPMLERIEETTKLGF-F 355  
 Db 676 ENLLEEVEKAVIADAEVKLQEKDKRCQKIAEMVALMEXKHQYDKYDIERDSELGLY 735  
 QY 356 KHPVIGVHVRTDKVGTAAAFHPHEEYVMVHEEPOLLERRMKVDKKRYLATDDPSLLK 415  
 Db 736 KSK---EQEQSSLASLELSNLKAEILLSVKKQLE-IEREEKEKLKRE--AKENTATLIK 789  
 QY 416 EAKTKSYNVEFISDINSISM 434  
 Db 790 EKKOKKKTOTFLLEPFIYW 808  
 RESULT 15  
 ID REST\_HUMAN STANDARD; PRT; 1427 AA.  
 AC P30622;  
 DT 01-APR-1993 (Rel. 25, (reated)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-Sternberg intermediate filament associated protein).  
RSN.

Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]

SEQUENCE FROM N.A.  
TISSUE=Peripheral blood monocytes;  
MEDLINE=92289675; PubMed=1600942;  
Bilbe G., Delabie J., Brueggen J., Richener H., Aseelbergs F.A.M., Cerletti N., Sorg C., Odink K., Tarcsey L., Wiesendanger W., de Wolf-Peters C., Shipman R.;  
"Restin: a novel intermediate filament-associated protein highly expressed in the Reed-Sternberg cells of Hodgkin's disease.";  
EMBO J. 11:2103-2113(1992).  
[2]

SEQUENCE FROM N.A.  
MEDLINE=92405160; PubMed=1356075;  
Pierre P., Scheel J., Rickard J.E., Kreis T.E.;  
"CLIP-170 links endocytic vesicles to microtubules.";  
Cell 70:887-900(1992).

-!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.  
-!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE CYTOSKELETON.  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=Long;  
IsoId=P30622-1; Sequence=Displayed;  
Name=Short;  
IsoId=P30622-2; Sequence=VSP\_000765;  
-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS OF HODGKIN'S DISEASE.  
-!- SIMILARITY: Contains 2 CAP-Gly domains.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X64838; CAA46050.1; -  
EMBL; M97501; AAA35693.1; -  
PIR; S22695; S22695.  
Genew; HGNC:10461; RSN.  
MIM; 179838; -  
GO; GO:0005768; C:endosome; TAS.  
GO; GO:0005882; C:intermediate filament; TAS.  
GO; GO:0015630; C:microtubule cytoskeleton; TAS.  
GO; GO:0008017; F:microtubule binding activity; TAS.  
GO; GO:0005899; P:non-selective vesicle transport; TAS.  
InterPro; IPR000938; CAP-Gly.  
InterPro; IPR001878; Znf\_CCHC.  
Pfam; PF01302; CAP GLY; 2.  
SMART; SM00343; Znf\_C2HC; 1.  
PROSITE; PS00845; CAP GLY 1; 2.  
PROSITE; PS50245; CAP GLY 2; 2.  
Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.  
DOMAIN 78 120 CAP-Gly 1.  
DOMAIN 143 204 SER-RICH.  
FTT DOMAIN 232 274 CAP-Gly 2.  
FTT DOMAIN 304 331 SER-RICH.  
FTT DOMAIN 350 1342 COILED COIL (POTENTIAL).  
DOMAIN 1408 1421 CCHC-BOX.  
FTT VARSPLIC 457 491 Missing (in isoform Short).  
CONFLICT 1069 1069 /FTIG=VSP\_000765.  
SEQUENCE 1427 AA; 160989 MW; 0AA4F166DD94254B8 CRC64;  
SQ

[illegible]

Search completed: February 2, 2004, 08:40:18  
Job time : 14.5 secs

CONFLICT SEQUENCE	1069 1427	1069 AA;	1069 MW;	D -> E (IN REF. 2).	0A4F166DD94254E8 CRC64;
SEQUENCE	1069	1069	1069	D -> E (IN REF. 2).	0A4F166DD94254E8 CRC64;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 33.5 Seconds  
(without alignments)  
4429.261 Million cell updates/sec

Title: US-09-971-773-23  
Perfect score: 3081  
Sequence: 1 MRWGTGSRWMTLILFANGT.....YKREKJETVKYPTYPEAK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*  
1: sp\_archea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2783.5	90.3	578	13 Q8AXS7	Q8axs7 xenopus lae
2	2176	70.6	446	4 Q8NEP2	Q8nep2 homo sapien
3	1489.5	48.3	619	5 Q9YVY5	Q9yvy5 d putative
4	991	32.2	559	5 O16882	O16882 caenorhabdi
5	810	26.3	169	4 Q8IUA5	Q8iua5 homo sapien
6	671	21.8	512	5 Q8IFW9	Q8ifw9 ciona intes
7	614.5	19.9	560	5 Q8IFW8	Q8ifw8 ciona intes
8	137.5	4.5	1137	16 Q8F3E7	Q8f3e7 leptospira
9	132.5	4.3	876	17 Q8TX14	Q8tx14 methanopyru
10	128	4.2	674	5 Q18106	Q18106 caenorhabdi
11	127.5	4.1	1218	5 Q8IED2	Q8ied2 plasmodium
12	124	4.0	1025	10 Q9SAA5	Q9saas arabidopsis
13	124	4.0	1285	4 Q9UEG2	Q9ueg2 homo sapien
14	124	4.0	2442	4 O14812	O14812 homo sapien
15	122.5	4.0	681	11 Q8BU18	Q8bul18 mus musculu
16	122.5	4.0	1200	11 Q921B9	Q921b9 mus musculu

17	122	4.0	1871	5 Q8IR54	Q8ir54 drosophila
18	122	4.0	2328	5 Q9VY43	Q9vy43 drosophila
19	122	4.0	2360	5 Q8IR55	Q8ir55 drosophila
20	122	4.0	2501	5 Q9NCW7	Q9ncw7 drosophila
21	121.5	3.9	975	13 Q98TQ5	Q98tc5 notothenia
22	121.5	3.9	3187	11 Q63714	Q63714 rattus norv
23	121	3.9	1285	4 Q9BZZ7	Q9bzz7 homo sapien
24	120.5	3.9	584	17 Q96YQ6	Q96yq6 sulfolobus
25	120.5	3.9	3542	5 Q9USM2	Q9usm2 plasmodium
26	119.5	3.9	612	4 Q96JV2	Q96jv2 homo sapien
27	119.5	3.9	2007	13 Q02015	Q02015 gallus gall
28	119.5	3.9	3616	13 Q9W6V0	Q9w6v0 gallus gall
29	119	3.9	1871	10 Q9SRD5	Q9srd5 arabidopsis
30	118	3.8	581	5 Q21065	Q21065 caenorhabdi
31	118	3.8	717	4 Q96ED9	Q96ed9 homo sapien
32	118	3.8	1999	11 Q63731	Q63731 rattus norv
33	117.5	3.8	1265	3 Q59920	Q59920 pneumocysti
34	117.5	3.8	1397	10 Q9LP90	Q9lp90 arabidopsis
35	117.5	3.8	1929	13 Q98TQ6	Q98tc6 notothenia
36	117	3.8	1177	16 Q8RCY8	Q8rcy8 thermoanaer
37	116	3.8	178	4 Q8WX78	Q8wx78 homo sapien
38	116	3.8	214	6 Q8MJ49	Q8mj49 sus scrofa
39	116	3.8	217	4 Q96I04	Q96ij4 homo sapien
40	116	3.8	283	10 Q9C717	Q9c717 arabidopsis
41	116	3.8	1156	16 Q66878	Q66878 aquifex aeo
42	116	3.8	1871	5 Q9NCL3	Q9nc13 drosophila
43	116	3.8	2442	4 Q9H450	Q9h450 homo sapien
44	116	3.8	2442	4 Q60588	Q60588 homo sapien
45	115.5	3.7	716	11 Q8BY47	Q8by47 mus musculu

#### ALIGNMENTS

#### RESULT 1

Q8AXS7 PRELIMINARY; PRT; 578 AA.  
AC O8AXS7;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative alpha-6-fucosyltransferase (EC 2.4.1.68).  
GN FUT8.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mollicone R., Michalski J.C., Bauvy C., Cailleau-Thomas A., Oriol R.;  
RA Candelier J.J., Martinez-Duncker I., Breton C., Codogno P., Oriol R.;  
RT "Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis."  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ514872; CAD55833.1; --  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 578 AA; 66578 MW; 2741A03BDB3C5265 CRC64;

Query Match 90.3%; Score 2783.5; DB 13; Length 578;  
Best Local Similarity 87.7%; Pred. No. 2e-192;  
Matches 507; Conservative 40; Mismatches 28; Indels 3; Gaps 1;

Qy	1	MRWGTGSRWMTLILFANGTLLFYIGGHLVRDNDPHSSRELSKIILAKLERLQKQNDL	60
Db	1	MRPWTGSRWMTLILFANGTLLFYIGGHLVRDNDPHSSRELSKIILAKLERLQKQNDL	60
Qy	61	RRMAESIRIPEGFTDOGTATGRVRLVEQLVKAKEQLENYKQARN---DLGKDEILRR	117
Db	61	RRMAESIRIPEGFTDOGTATGRVRLVEQLVKAKEQLENYKQARN---DLGKDEILRR	117
Qy	118	RIENGAKELWFFIIOSELKKLKLKLEGNLQRLHDEILLDLGHHERSIMTDLYLSSQTDGAG	177
Db	118	RIENGAKELWFFIIOSELKKLKLKLEGNLQRLHDEILLDLGHHERSIMTDLYLSSQTDGAG	177

```

Db 121 AIEGAKFVYLVQSEVKKLKHLDRLNQLRHVDEILIDMGHQRSVMTDLYLSTQDAG 180
Qy 178 EWEKEAKDLTELVRRTYILQNPCKSKARKLVNINKGCGYQGLHHVYCFMAYGT 237
Db 181 DWREAEKDLTLVQRRTYILQNPCKSKAKKLVNINKGCGYQGLHHVYCFMAYGT 240
Qy 238 QRTLLSQNRWYATGGHETFRPVSETCTDRSGLSTGHWGEVCKNVQVPELPIVDSL 297
Db 241 QRTLLSQNRWYATGGHETFRPVSETCTDRSGLSTGHWGEVCKNVQVPELPIVDSL 300
Qy 298 HPRPPLPLAVPEDLADLRLLRHGDPVAVWVSQFVKYLIRPQWLERIEETTKLGFPH 357
Db 301 HPRPPLPLAVPEDLADLRLLRHGDPVAVWVSQFVKYLIRPQWLERIEETTKLGFPH 360
Qy 358 PVIGVHVRTDKVGTAAFPHPPIEYVWVHEHFQLLERMMKVDKRVYLATDDPSLLKEA 417
Db 361 PVIGVHVRTDKVGTAAFPHPPIEYVWVHEHFQLLERMMKVDKRVYLATDDPSLLKEA 420
Qy 418 KTKYNYEFISDNSISWSAGLHNRYTENSLRGVLDIHFLSQADFLVCTFSSQVCRVAYE 477
Db 421 KAKYQYEFISDNSISWSAGLHNRYTENSLRGVLDIHFLSQANFLVCTFSSQVCRVAYE 480
Qy 478 IMQTLHPDASANFHSLLDDIYFGGQNAHNOIAVPHQPRTEKEIIPMEPCDIIIGVAGNHWN 537
Db 481 IMQTLHPDASANFHSLLDDIYFGGQNAHNOIAVPHQPRTEKEIIPMEPCDIIIGVAGNHWN 540
Qy 538 GYSKGVNRKLGKTLGTPSYKVKREKTIETVKYPTYPEAEK 575
Db 541 GYSKGVNRKLGKTLGTPSYKVKREKTIETVKYPTYPEAEK 578

RESULT 2
Q8NEP2
ID Q8NEP2 PRELIMINARY; PRT; 446 AA.
AC Q8NEP2;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative alpha-(1,6)-fucosyltransferase (GDP-fucose-glycoprotein
DE alpha-L-fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUCT6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RT Petit D., Picaud F., Dupuy F., Germot A., Julien R., Maftah A.;
RT "Core a3- and a6-fucosyltransferases in Drosophila: characterization
RT and origin of diversity.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwa C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

```

```

Qy 401 KKRVLATDDPSLLKEATKYSNYEFISDNSISWSAGLHNRYTENSLRGVLDIHFLSQ 460
Db 272 KKRVLATDDPSLLKEATKYPNYEFISDNSISWSAGLHNRYTENSLRGVLDIHFLSQ 331
Qy 461 DFLVCTFSSQVCRVAYEIMOTLHPDASANFHSLLDDIYFGGQNAHNOIAVPHQPRTEKE 520
Db 332 DFLVCTFSSQVCRVAYEIMOTLHPDASANFHSLLDDIYFGGQNAHNOIAVPHQPRTEKE 391
Qy 521 IPMEPCDIIIGVAGNHWNYSKGVNRKLGKTLGTPSYKVKREKTIETVKYPTYPEAEK 575
Db 392 IPMEPCDIIIGVAGNHWNYSKGVNRKLGKTLGTPSYKVKREKTIETVKYPTYPEAEK 446

RESULT 3
Q9VVV5
ID Q9VVV5 PRELIMINARY; PRT; 619 AA.
AC Q9VVV5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-
DE alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUCT6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RT Petit D., Picaud F., Dupuy F., Germot A., Julien R., Maftah A.;
RT "Core a3- and a6-fucosyltransferases in Drosophila: characterization
RT and origin of diversity.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwa C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

```





```
DR EMBL; AJ512486; CAD54736.1; -.
KW Hypothetical protein; Glycosyltransferase; Transferase.
SQ SEQUENCE 559 AA; 63648 MW; BB6F6F1944A9B9F9F CRC64;

Query Match      32.2%; Score 991; DB 5; Length 559;
Best Local Similarity 36.7%; Pred. No. 4.1e-63;
Matches 213; Conservative 115; Mismatches 188; Indels 64; Gaps 11;

QY 3 AWTGSRWIMLILFWAGTLLFYIGGHLVRDNDPHDSSRELKILAKLERLKQOQNEIDLRR 62
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6 AAVGTVMWMTFLYLSQ-----SNNQSGDSIRAWRQTKEADIKLQEQNEIDLKS 56

QY 63 MAESRIPEGPTDQGTATGRVLEBQVKAQEOTENYKKQARN-----107
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
57 ILEKER-----QERNDQHKIMEQSHQLPPNPENSLPKPEPVKEIISK 101

QY 108 --LGR-DHEILLRRLENGAKELWFFLOSELKLLKLEGNELQORHADEILLDGHHERSIM 164
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
102 SILGPVQVQVRKMLDDRIEMFYLLHSQTIE-----NSTKILLET--QMSILM 148

QY 165 TDLVYLSQTDGAGWEKEAKDLTELVRIRITYLQNPDKCSKARKLVNINKGCGYGCQL 224
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 GLSAQLEKLEGESEERFKQRTAITQRIKPSIEKLQNPACSEAKTLVNCNLDKECGFGCQL 208

QY 225 HHVYCFMAYQTQTLIL--ESQNRVATGWETVFRPVSETCTDRSGLSTGHSWGE-- 280
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 HHVTCATATAPATQRMVLRKDGSSWYSSHGWTSVFKKLSKCSFDE---AVGNTEAKPF 265

QY 281 VKDKNVQVVELIVDSLSLPRPYLPLAVPELDLRLVHGDPAVWVYSQFKYLIRPQP 340
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
266 AEPSPARVSLGIVDSLITKTFPLPOAPEQLLESLSLHSHPPAFFVGTFTISYLMRFS 325

QY 341 WLREIEETTKKLGK-KHPVIGVHVRRTDKVGTAAFPPIEYVMVHVEEHFOLLERRM-K 398
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 ATQEKLDKALKSIPLDKGPVGLQIRTDKVGTEAFAFHALKEYMEWTEIWFKEKROGK 385

QY 399 VDKRVYATDDPSLLKKAETKYSNYEFISDINSWSAGLHNRYTENSILRGVILDIHPLS 458
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 PLERRIFTASDDPTVPEAKNDYPNVYVGVSTPEIAKTAQLANNRYTASLIMGVITDIYLS 445

QY 459 QADFLVCTFSSQCVAVVEIMQTLHPDASANPHSLDDIYFGGQNAHNQIAVYVPHQPTK 518
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
446 KNYILVCTFSSQCVRMGLRPSGADGSKFSLDDIYFGGQQAHEVIVIEDHIAQNN 505

QY 519 BEIPMEPGDIIAGVAGNHNGYSKGVNRKLGKTGLYPSYKV 558
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
506 KEIDLKVGDKVGIAGNHNGYSKGVNRQTYKEGVFSPYKV 545

RESULT 5
Q8IUAS PRELIMINARY; PRT; 169 AA.
AC Q8IUAS;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Fucosyltransferase 8.
FUT8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22326102; PubMed=12438718;
RA Coullin P., Crocijsmans R.P.M.A., Groenen M.A.M., Heilig R.,
RA Mollicone R., Oriol R., Candelieri J.J.;
RT "Assignment of FUT8 gene to chicken chromosome band 5q1.4 and to human
RT chromosome 14q23.2-24.1 by in situ hybridization. Conserved and
RT compared synteny between human and chicken.";
RL Cytogenet. Genome Res. 97:234-238(2002).
DR EMBL; AJ514324; CAD55804.1; -.

DR EMBL; AJ514325; CAD55805.1; -.
KW Transferase.
SQ SEQUENCE 169 AA; 19520 MW; 0ACD43AA0F56AB31 CRC64;

Query Match      26.3%; Score 810; DB 4; Length 169;
Best Local Similarity 94.9%; Pred. No. 9.2e-51;
Matches 148; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 420 KYSNYEFISDINSWSAGLHNRYTENSILRGVILDIHFLSQADFLVCTFSSQCVAVYBIM 479
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 RYPNYEFISDINSWSAGLHNRYTENSILRGVILDIHFLSQADFLVCTFSSQCVAVYBIM 73

QY 480 QTLHPDASANPHSLDDIYFGGQNAHNQIAVYVPHQPTKEIIPMEPGDIIAGVAGNHNGY 539
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
74 QTLHPDASANPHSLDDIYFGGQNAHNQIAVYVPHQPTADEIIPMEPGDIIAGVAGNHNGY 133

QY 540 SKGVNRKLGKTGLYPSYKVKREKIETVYKPYTPYPAEK 575
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 SKGVNRKLGKTGLYPSYKVKREKIETVYKPYTPYPAEK 169

RESULT 6
Q8IFW9 PRELIMINARY; PRT; 512 AA.
AC Q8IFW9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative alpha-6-fucosyltransferase (EC 2.4.1.68).
FUT8.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]_
RP SEQUENCE FROM N.A.
RA Mollicone R., Michalski J.C., Bauvy C., Cailleau-Thomas A., Oriol R.;
RA Candelieri J.J., Martinez-Duncker I., Breton C., Codogno P., Oriol R.;
RT "Splice variants of alpha-6-fucosyltransferase are expressed early in
RT human embryogenesis.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ515151; CAD56161.1; -.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 512 AA; 59217 MW; 85FC7F702EC58178 CRC64;

Query Match      21.8%; Score 671; DB 5; Length 512;
Best Local Similarity 30.1%; Pred. No. 4.4e-40;
Matches 157; Conservative 93; Mismatches 200; Indels 72; Gaps 13;

QY 82 RRVLESQLVKAQEIQENYKKQARNDLGKHDEILRRRIENGAKELWFFLOSELKLLKLE 141
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
21 KVRSLRSIVKSPQIQQLLNHDMNKLAKNKLQVQLVN-----W----- 61

QY 142 GNEQLORHADEILLDGHHERSIMTDLVYLSQTDGAGWEKEAKDLTELVRIRITYLQNP 201
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 -----DRLVPL-----ENSLINEMTLLNNKDEGD--RYKRELNRRAHQLIQTQNO 108

QY 202 KDCSKARKLVNINKGCGYCOLHHVYVCFMAYGTORTLILESQNRVAT-----GGMET 257
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
109 KNCERAKLFCNVEAKCGFGLVHHYGVCLFLAIGTRVM-----QWDLKAITDYPGLDE 163

QY 258 VFRPEVSETCTD-RSGLSTGHSWGEVKDN-----VQVVELPVDLSLHPRPYLPLAVPED 311
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 VFMPLESETCNITGCKEAPWYS-KSNSPTPVDTEVVKVTLIDSHDRQTFPAMTIPPE 222

QY 312 LADLLRVHGDPAVWVYSQFKYLIRPQPLERIEETTKKLGFKHPVGVHVRTKVG 371
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
223 FYPLVKPIHGDPVWVWAGVQSYLMRPAWVTRVQIIKQDIGFKHPVGVHVRTK-S 281

QY 372 TEAAFPHEEYVWVHEBFHQ--LIERRMKVDKRVYLATDDPSLLKKAETKYSNYEFISD 429
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 NEANFORMEAYMDPVANWYDKYVINDVTNITKLIVYVATDDLEVMYDMLRNLYPEFTITN 341
```





Qy 399 VDKRRVLTDDPSLLKAKTKYNGYEFISDINSISWSAGLHRYTEN 445  
Db 626 AVEKRTAAKSLDMSLRKTE-----IDYEDIKTESIRLQEEYNNMLLTN 668

## RESULT 11

Q8IED2 PRELIMINARY; PRT; 1218 AA.  
AC Q8IED2  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Chromosome segregation protein, putative.  
GN MAL13P1.96.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Harris B., Lennan N., Clark L., Line A., Barron A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52328.1;  
SQ SEQUENCE 1218 AA; 143059 MW; E1DBD7838BBC7E5B CRC64;

Query Match 4.1%; Score 127.5; DB 5; Length 1218;

Best Local Similarity 24.8%; Pred. No. 2.4;  
Matches 60; Conservative 41; Mismatches 52; Indels 89; Gaps 14;

Qy 41 RELSKILAKLER-----LQONEDLRRMAESLRPEPIDQGTATGRVVRVLEEQIVK 92  
Db 784 KKLTEVIRKLEKIDSEYANKDKKEEDLKEIKLK-----NKKIQLETEEHK 831  
Qy 93 AKE-----QIENYKQAR---NDLGKDHILR-----RRIENG 122  
Db 832 KKEEIDVLLQIENYKQKKEETNDLSSTDEIINEIEKKIEDIEKNINITRENKLEENK 891  
Qy 123 AKEL---WFFQSELEK-LKLEGNELQHADEILDLGHHSRTMTDLY----- 168  
Db 892 ITELQSSFSSEYENEMKHVVKKIE--DLEKKKSENILDLKULENTLL-DLQKDLKTSSTV 948  
Qy 169 -YLSOTDGCAGWEKEA-----KDLT-----ELVORRITYLQNPDCSKARKLVGNI 214  
Db 949 KYLYKTH---VWIESYELPNKNTPTYPDFENFRHDVIOKKIQALQNEQN-----KLSINI 1000  
Qy 215 NK 216  
Db 1001 NR 1002

## RESULT 12

Q9SAA5 PRELIMINARY; PRT; 1025 AA.  
AC Q9SAA5  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE F25C20.13 protein.  
GN F25C20.13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Lee J.M., Kremenetskaia I., Lueros J., Ngan I., Liu A.,  
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,  
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,  
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;

RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Theologis A.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC007296; AAD30351.1; -

SQ SEQUENCE 1025 AA; 116411 MW; 08C952A2032BA1E4 CRC64;

Query Match 4.0%; Score 124; DB 10; Length 1025;

Best Local Similarity 18.6%; Pred. No. 3.4;

Matches 115; Conservative 79; Mismatches 184; Indels 240; Gaps 32;

Qy 5 TGSWRWIMLILFAWGTLLP-YI-----GHLVRDNDHPDHSRRELSKILAKLERLKKQ 56  
Db 362 SGNW-WFAEVVVGALVIDWVADGPPKGAFLYDNGYQDFHALVPQKLPPELYWLEEE 420  
Qy 57 NEDLRRMAESLRPEGP-----DQGTATGRVVRVLEEQIVKAKQEIENYKQARNDLG 109  
Db 421 NMIFRKLEDRRLKEEVYRAKMEKTARLKAETKERTLKLLSKQVW--YTEPLEIQAG 478  
Qy 110 KDHEIL---RRRTEGAKELWF--FLOSELKLLKLEGNELQHADE-----ILLD 155  
Db 479 NPVTLYNPANTVILNGKPEVWFRGSRNWRTHRLGPLPPQKMEATDDESSHVKTAKVPLD 538  
Qy 156 LGHHSRIMTDLVYLSOTDGCAGWEKE-----AK 185  
Db 539 -----AYMDFVFEKEDG-GIFDNKNGLDYHLPVVGIGISKEPPLHIVHIAVEMAPIAK 591  
Qy 186 ---DLTELVRRIITYLQN-----PK-DCSK---ARKLVN----- 213  
Db 592 VGGIGDVVTSLSRAVQELNHNVDIVFPKYDCIKHNFVKDLQFNRSYHWGTEIKVWHGKV 651  
Qy 214 -----INKGCGVGCQ-----LHHVYCFMIAYGTOFTLILESQNRWYA 251  
Db 652 EGLSVYFLDPQNLFGRCVYGCADDAGRFQFFCHAALEFLLOGGFHPD--ILHCHDWSSA 710  
Qy 252 TGGWETVFRPVSITCTDRSGLSGTGHS--GEVQKQVQVWELPIVDSLH----- 298  
Db 711 PVSW-----LFKDHVYQYGLIKTR-----IVFTIHNLEFGANAIGK 746  
Qy 299 -----PRPYL-PLAVPEDLADRLRLRVHG-----DPAVW--WVSQVVKYLIIRQP 340  
Db 747 AMTFADKATTVSPYAKEVAGNSVISAHLYKFHGIINGIDPDIWDPDYNDNFI-----PVP 801  
Qy 341 WLREI-----EETTKLGLFK---HPVIGVHVRRTDKVGTAAAPHPIEYVWVVEEH 389  
Db 802 YTSENVVVEGKRAAKEELQNLRLGLKSADFPVVGIIITRLTHQKG-----IHLIKH 849  
Qy 390 --FOLLERRMKV-----DKKRVYLATDDPSLLKEAKTK 420  
Db 850 AIWETLERNQGVLLGSAPDPRIQDNFVNLANQLHSHSGDRARLVITYDEPL-----SHLI 905  
Qy 421 YSNYEFTSDNSIWSAGL 438  
Db 906 YAGADFILVPSIPEPCL 923

## RESULT 13

Q9UEG2 PRELIMINARY; PRT; 1285 AA.  
AC Q9UEG2  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein KIAA0389.  
GN KIAA0389.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]



```

Qy 23 FYIGHLVRDNDHPSRSLKILAKLERLKQONEDLRMAESLRIPGEPIDQGTATGR 82
Db 55 FEIEKRLSQSERLVTTRECONLRLELEKLNQVKVLTKEKLE-----TAQDR 105
Qy 83 VRLEEQVLKAKEQIENYKQARDNLGKHDEILRRRIENGAKELWFFLOSELKKL--KKL 140
Db 106 NLGIQSOFTRAKELEAEKR---DLIRTNERLSQVEB-----YLTEDVKRLNEKJK 153
Qy 141 EGN---ELORHAEIILLDGHHERSINTDLYILSQTDTGAGEWEKEAKDLTELVRORIT 196
Db 154 ESNTTKGELQKLDE-----LOADSVAVKYREKRLQEKEKELLHNQNS 195
Qy 197 YLQNPDKCSKARKLVCNINKG---CGYGCQLHHVYVCFMIAYGTOFTLILESQNRYATG 253
Db 196 WLNTLTKTKTDDELLALGREKGNELILELKNLEN-----KKEEVLRLLEEQWNLKTS 246
Qy 254 GWETVFRPVSETCTDRSLSTGHSGEVKQKNQVVELPI-----VDSLHPPPYLPL 306
Db 247 N-EHLQKHVEDLLTKLKEAQASMEEFHNLNAHIKLSNLYKSAADDSEAKSNELTR 305
Qy 307 AVPE-----DLADLLRVHGDPAVWVVSQFVKYLIRPQWLEREIEETTKL 353
Db 306 AVDELHKLKKEAGEANKTIQDHLLOVESK-----DQMEKEMLEKIGKLEKELENANDLL 360
Qy 354 GFKHPVIGVHVRRTDKVQTE---AAFHP-----IEEYMHVVEEHFOLLER 395
Db 361 S-----ATYKRGAILSEELAAWSPATAAAKIVKPGMKLTLYNAYVETODQLLLE 412
Qy 396 RMKVDKRVYLATDDPSLLKEAKTKYSNYEFISDNISWSAGLHNRYTENSIRGVILDIH 455
Db 413 --KQENKRIKYLDE--IVKEVEAK-----APILKQOREE-----443
Qy 456 FLSQADFLVCTFSQVCRVAYEIMOTLHPDASANFHS 492
Db 444 -YERAKAVASLSAKLEQAMKEIQRLQEDTDKANKHS 479

```

Search completed: February 2, 2004, 08:43:08  
Job time : 37.5 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 38.5 Seconds  
(without alignments)  
2370.592 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRAWTGSWRWMLILFAWGT.....YKREKIETVKYTPPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03:\*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3085	100.0	575	23 ABG34136	Antibody productio
2	3024	98.0	575	23 ABG34135	Antibody productio
3	3003	97.3	575	18 AAW22125	Human alpha 1-6 fu
4	3003	97.3	575	23 ABB08405	Alpha1,6-fucosyl t
5	2910	94.3	575	18 AAW22124	Pig alpha 1-6 fuco
6	2658	86.2	515	22 AAG73884	Human colon cancer
7	1778	57.6	339	22 AAB75061	Human alpha 1-6 fu
8	1480.5	48.0	619	22 ABB59117	Drosophila melanog
9	1206	39.1	233	22 AAB75062	Human alpha 1-6 fu

10	432	14.0	82	22	ABG48804	Human liver peptid
11	432	14.0	82	22	ABB28797	Peptide #1448 enco
12	432	14.0	82	22	ABB33985	Peptide #1491 enco
13	432	14.0	82	22	ABB19422	Protein #1421 enco
14	432	14.0	82	22	AAW54747	Human brain expres
15	432	14.0	82	22	AAW67139	Human bone marrow
16	432	14.0	82	22	AAW15003	Peptide #1437 enco
17	432	14.0	82	22	AAW27440	Peptide #1477 enco
18	432	14.0	82	22	AAW02730	Peptide #1412 enco
19	432	14.0	82	22	ABG36801	Human peptide enco
20	426	13.8	81	22	AAW64859	Human brain expres
21	426	13.8	81	22	AAW37789	Peptide #1026 enc
22	131	4.2	25	18	AAW22131	Fragment #2 of hum
23	125	4.1	1025	23	ABB90967	Herbicidally activ
24	120.5	3.9	751	22	ABB62477	Drosophila melanog
25	120.5	3.9	3542	22	AAW62142	P. falciparum PCR3
26	118.5	3.8	976	24	ABF74709	Human SCP-1 protei
27	117	3.8	1456	22	ABB58673	Drosophila melanog
28	117	3.8	1521	21	AAG39235	Arabidopsis thalia
29	117	3.8	1528	21	AAG39234	Arabidopsis thalia
30	117	3.8	1552	21	AAG39233	Arabidopsis thalia
31	117	3.8	1703	21	AAG36714	Arabidopsis thalia
32	117	3.8	1710	21	AAG36713	Arabidopsis thalia
33	117	3.8	1744	21	AAG36712	Arabidopsis thalia
34	116.5	3.8	795	23	ABB77430	Human tumour marke
35	116	3.8	1235	24	ABR47540	Breast cancer asso
36	115.5	3.7	931	22	AAW79504	Human protein SEQ
37	115.5	3.7	930	22	AAW78520	Human protein sequ
38	114.5	3.7	612	22	AAW95466	Human homologue of
39	114.5	3.7	888	23	AAU83013	Human 160kD mediat
40	114.5	3.7	1427	12	AAW10534	Osteoclast stimula
41	114	3.7	214	18	AAW09037	Osteoclast stimula
42	114	3.7	214	18	AAW12706	Osteoclast stimula
43	114	3.7	229	17	AAW05403	Human clone 5 prot
44	113.5	3.7	976	22	AAG66581	Human SCP-1 mucin
45	113	3.7	26	18	AAW22126	Fragment #1 of Pig

#### ALIGNMENTS

#### RESULT 1

ABG34136  
ID ABG34136 standard; Protein; 575 AA.  
AC ABG34136;  
XX  
XX  
15-JUL-2002 (first entry)  
DT  
XX  
XX  
Antibody production method related protein #1.  
DE  
XX  
Antibody production; cytostatic; immunomodulator; vasotropic; virucide;  
KW  
antibacterial; antiinflammatory; antiallergic; allergy; inflammation;  
KW  
autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;  
KW  
tumour; circulatory disease; infection.  
XX  
XX  
Mus musculus.  
OS  
XX  
WO200231140-A1.  
FN  
XX  
18-APR-2002.  
PD  
XX  
05-OCT-2001; 2001WO-JF08804.  
PF  
XX  
06-OCT-2000; 2000JP-0308526.  
PR  
XX  
(KYOW ) KYOWA HAKKO KCGYO KK.  
PA  
XX  
Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;  
PI  
Hosaka E, Yamano K, Yamasaki M, Hanai N;  
XX  
WPI; 2002-340182/37.  
XX





```
PR 22-MAR-2000; 2000JP-0081059.
XX
XX (TANI/) TANIGUCHI N.
PA (SEKI/) SEKI T.
PA (FUJ/) FUJIYAMA K.
XX
XX WPI; 2002-159816/21.
DR N-PSDB; ABA98809.
XX
XX A plant cell with an animal type sugar chain adding function, for the
PT preparation of a glycoprotein with an animal type sugar chain -
XX
XX Example 1; Page 24-26; 38pp; Japanese.
XX
XX The invention relates to a plant cell with an animal type sugar chain
CC adding function, created by transforming a tobacco cell with a gene
CC encoding an enzyme derived from an animal which can transfer a fucose
CC residue to the reductive end acetylglucosamine residue of a sugar chain.
CC The gene that is introduced into the plant cell encodes the enzyme
CC alpha1,6-fucosyl transferase. The method of the invention is useful for
CC the preparation of a glycoprotein having animal type sugar chain. The
CC current sequence represents alpha1,6-fucosyl transferase amino acid
CC sequence.
XX
XX Sequence 575 AA;
SQ
Query Match 97.3%; Score 3003; DB 23; Length 575;
Best Local Similarity 96.5%; Pred. No. 1.4e-271;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
QY 1 MRWGTGSRWIMLILFWAGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
Db 1 MRWGTGSRWIMLILFWAGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPGPIDQGTATGRVRLEROLVKAKEQIENYKQANGLGKDHILRRRIE 120
Db 61 RMAESLRIPGPIDQGTATGRVRLEROLVKAKEQIENYKQANGLGKDHILRRRIE 120
QY 121 NGAKELWFFLQSELKHLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180
Db 121 NGAKELWFFLQSELKHLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180
QY 181 EKEAKDLTELVRRTIYQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMIAYGTQRT 240
Db 181 EKEAKDLTELVRRTIYQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMIAYGTQRT 240
QY 241 LILESQNRWYATGGWETVPRVSETCTDRSGISTGHSGEVNDKNIQVVELPIVDSLHPR 300
Db 241 LILESQNRWYATGGWETVPRVSETCTDRSGISTGHSGEVNDKNIQVVELPIVDSLHPR 300
QY 301 PPYLEPLAVPEDLADRLRLVHGDPVWVWSQFVKYLIRPQWLEKEIEBATKLGPKHPVI 360
Db 301 PPYLEPLAVPEDLADRLRLVHGDPVWVWSQFVKYLIRPQWLEKEIEBATKLGPKHPVI 360
QY 361 GVHVRTDKVGTAAFPPIEYVWVHEHFQLLARMQVQDKRVYLATDDPTLLKEATK 420
Db 361 GVHVRTDKVGTAAFPPIEYVWVHEHFQLLARMQVQDKRVYLATDDPTLLKEATK 420
QY 421 YSNYEFISDINSISWAGLHNRYTENSRLGVLDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
Db 421 YSNYEFISDINSISWAGLHNRYTENSRLGVLDIHFLSQADFLVCTFSQVCRVAYEIMQ 480
QY 481 TLHPDASANFHLDDIYYPGGQANNOIAVYHPKPTREEIPMEPGDIIIGVAGNHWDCYS 540
Db 481 TLHPDASANFHLDDIYYPGGQANNOIAVYHPKPTREEIPMEPGDIIIGVAGNHWDCYS 540
QY 541 KGINKRLGKTGLYPSYKVRKTIETVKYPTYPEAK 575
Db 541 KGINKRLGKTGLYPSYKVRKTIETVKYPTYPEAK 575
RESULT 5
AAW22124
```

```
AAW22124 standard; Protein; 575 AA.
AAW22124;
05-MAR-1998 (first entry)
Pig alpha 1-6 fucosyltransferase.
Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;
guanosine diphosphate; sugar chain synthesis; modification; antibody;
GlcNAc; cancer diagnosis.
Sus scrofa.
W09727303-A1.
31-JUL-1997.
23-JAN-1997; 97WO-JP00171.
22-JUL-1996; 96JP-0192260.
24-JAN-1996; 96JP-010365.
21-JUN-1996; 96JP-0161648.
24-JUN-1996; 96JP-0162813.
(TOYM) TOYO BOSEKI KK.
Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
WPI; 1997-393690/36.
N-PSDB; AAT76573.
Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for
synthesis and modification of sugar chains and used as an antigen
for production of diagnostic antibodies
Claim 4; Page 30-34; 61pp; Japanese.
AAW22124 and AAW22125 represent the pig and human alpha 1-6
fucosyltransferases of the invention, respectively. The enzyme transfers
fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc
nearest to R in the receptor molecule: (GlcNAc-beta 1-2Man-alpha 1-6)
(GlcNAc-beta 1-2Man-alpha 1-3)Man-beta 1-4GlcNAc-beta 1-4GlcNAc-R to give
(GlcNAc-beta 1-2Man-alpha 1-6) (GlcNAc-beta 1-2Man-alpha 1-3)Man-beta
1-4GlcNAc-beta 1-4 (Fuc-alpha 1-6)GlcNAc-R. It has an optimum pH of about
7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5
hours at 4 degrees C. The optimum working temperature of the
alpha 1-6 fucosyltransferases is 30-37 degrees C. A bivalent metal is
not required for activity of the enzyme, and the enzyme is not inhibited
in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
modification of sugar chains, and as antigen for the production of
antibodies recognising the enzyme. The antibodies can be used for the
diagnosis of cancer and other diseases.
Sequence 575 AA;
Query Match 94.3%; Score 2910; DB 18; Length 575;
Best Local Similarity 93.6%; Pred. No. 7.1e-263;
Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;
QY 1 MRWGTGSRWIMLILFWAGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
Db 1 MRWGTGSRWIMLILFWAGTLLFYIGGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
QY 61 RMAESLRIPGPIDQGTATGRVRLEROLVKAKEQIENYKQANGLGKDHILRRRIE 120
Db 61 RMAESLRIPGPIDQGTATGRVRLEROLVKAKEQIENYKQANGLGKDHILRRRIE 120
QY 121 NGAKELWFFLQSELKHLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180
Db 121 NGAKELWFFLQSELKHLKLEGNELQRADEFLDLGHHSRIMTDLVYLSQTDGAGDWR 180
QY 181 EKEAKDLTELVRRTIYQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMIAYGTQRT 240
```

Db 181 EKEAKDLTELQRRITLYLQNPDKCSKAKLVNKNKGGCGQLHHVVCYFMAVGTQRT 240  
Qy 241 LILESQNRVATGGWETFRPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPR 300  
Db 241 LALESHNNRYATGGWETFRPVSETCTDRSGSSTGHWSGEVNDKNIQVVELPIVDSLHPR 300  
Qy 301 PYPPLAVPEDLADRLVHVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGPKHPVI 360  
Db 301 PYPPLAVPEDLADRLVHVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGPKHPVI 360  
Qy 361 GVHVRTDKVTEAAFPHEEYVHVEHFFQLLARRMQVVKRVLATDDPTLLKEATK 420  
Db 361 GVHVRTDKVTEAAFPHEEYVHVEHFFQLLARRMQVVKRVLATDDPTLLKEATK 420  
Qy 421 YSNYEFISDNTSWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQ 480  
Db 421 YSNYEFISDNTSWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQ 480  
Qy 481 TLHPDASANFSLDDIYFVGGQNAHNOIAVPHKPRTEEEIPMEPGDIIGVAGNHWDCYS 540  
Db 481 ALHPDASANFSLDDIYFVGGQNAHNOIAVPHKPRTEEEIPMEPGDIIGVAGNHWDCYS 540  
Qy 541 KGINKLGTGLYPSYKREKIEVKYPTYPEAK 575  
Db 541 KGVNKRGTGLYPSYKREKIEVKYPTYPEADK 575

RESULT 6  
AAG73884  
ID AAG73884 standard; Protein; 515 AA.  
XX AC AAG73884;  
XX DT 03-SEP-2001 (first entry)  
XX DE Human colon cancer antigen protein SEQ ID NO:4648.  
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX KW colorectal carcinoma; chromosome 14.  
XX OS Homo sapiens.  
XX PN WO200122920-A2.  
XX PD 05-APR-2001.  
XX PF 28-SEP-2000; 2000WO-US26524.  
XX PR 29-SEP-1999; 99US-0157137.  
XX PR 03-NOV-1999; 99US-0163280.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX DR WPI; 2001-235357/24.  
XX DR N-PSDB; AAH33315.

NT nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
useful for preventing, diagnosing and/or treating colorectal cancers -  
Claim 11; Page 6451-6453; 9803pp; English.  
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
cancer-associated nucleic acid molecules (N) and proteins (P), where  
the proteins are collectively known as colon cancer antigens. The colon  
cancer antigens have cytostatic activity and can be used in gene  
therapy and vaccine production. N and P may be used in the prevention,  
diagnosis and treatment of diseases associated with inappropriate P  
expression. For example, N and P may be used to treat disorders  
associated with decreased expression by rectifying mutations or deletions  
in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX

Qy 65 ESLRIPGPDIDQ37ATGRVRLVLEBQLVKAKQIENYKQARGLGKHIEILRRRIENGAK 124  
Db 5 KSLRIPGPDIDQ37PAIGRVRLVLEBQLVKAKQIENYKQARGLGKHIEILRRRIENGAK 64  
Qy 125 ELWEFLOSELKK3KHLGNELOHRADELTLDLGHHSIMTDLYLSQDGCAGDWREKA 184  
Db 65 ELWEFLOSELKK3KHLGNELOHRADELTLDLGHHSIMTDLYLSQDGCAGDWREKA 124  
Qy 185 KDLTELQRRITLYLQNPDKCSKAKLVNKNKGGCGQLHHVVCYFMAVGTQRTTILE 244  
Db 125 KDLTELQRRITLYLQNPDKCSKAKLVNKNKGGCGQLHHVVCYFMAVGTQRTTILE 184  
Qy 245 SONRYATGGWETFRPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPRPPL 304  
Db 185 SONRYATGGWETFRPVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPRPPL 244  
Qy 305 PLAVPEDLADRLVHVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGPKHPVIGVHV 364  
Db 245 PLAVPEDLADRLVHVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGPKHPVIGVHV 304  
Qy 365 RRTDKVGTAAFPHEEYVHVEHFFQLLARRMQVVKRVLATDDPTLLKEATKYSNY 424  
Db 305 RRTDKVGTAAFPHEEYVHVEHFFQLLARRMQVVKRVLATDDPTLLKEATKYSNY 364  
Qy 425 EFISDNTSWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQTLHP 484  
Db 365 EFISDNTSWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYETMQTLHP 424  
Qy 485 DASANFSLDDIYFVGGQNAHNOIAVPHKPRTEEEIPMEPGDIIGVAGNHWDCYSKGIN 544  
Db 425 DASANFSLDDIYFVGGQNAHNOIAVPHKPRTEEEIPMEPGDIIGVAGNHWDCYSKGIN 544  
Qy 545 RKLGTGLYPSYKREKIEVKYPTYPEAK 575  
Db 485 RKLGTGLYPSYKREKIEVKYPTYPEAK 515

RESULT 7  
AAB75061  
ID AAB75061 standard; Protein; 339 AA.  
XX AC AAB75061;  
XX DT 20-JUL-2001 (first entry)  
XX DE Human alpha 1-6 fucosyltransferase protein 237-575 SEQ ID NO:3.  
XX KW Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;  
XX KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunoassay.  
XX OS Homo sapiens.  
XX PN JP20001011097-A.  
XX PD 16-JAN-2001.  
XX

```

PF 29-JUN-1999; 99JP-0183569.
XX
XX 29-JUN-1999; 99JP-0183569.
XX
XX (FJRE ) FUJIREBIO KK.
XX
XX WPI; 2001-275926/29.
XX
XX N-PSDB; AAF87952.
XX
XX Novel anti-human alpha1-6 fucose transferase antibody useful for
XX immunoassay -
XX
XX Example 2; Page 8-9; 1lpp; Japanese.
XX
XX The present invention describes an anti-human alpha 1-6
XX fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
XX transferase) antibody, also described are: (1) a hybridoma producing
XX the above monoclonal antibody; (2) an immunoassay for detecting human
XX alpha 1-6 fuct by using the above antibody or its antibody fragment;
XX and (3) a reagent used for the above immunoassay. The anti-human
XX alpha 1-6 fuct antibody can be used for immunoassay. The present
XX sequence represents the human alpha 1-6 fuct protein of residues 237
XX to 575, which is used in an example from the present invention.
XX
XX Sequence 339 AA;
SQ
Query Match 57.6%; Score 1778; DB 22; Length 339;
Best Local Similarity 96.2%; Pred. No. 2.3e-157;
Matches 326; Conservative 9; Mismatches 4; Indels 0; Gaps 0;
QY 237 TORTLLESQNRWYATGGWETVFRPVSETCTDRSLGSLTGHWSGEVNDKNIQVVELPIVDS 296
DB 1 TORTLLESQNRWYATGGWETVFRPVSETCTDRSLGSLTGHWSGEVNDKNIQVVELPIVDS 60
QY 297 LHPRPPLVLPVDELADRLVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFK 356
DB 61 LHPRPPLVLPVDELADRLVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFK 120
QY 357 HPVIGVHVARTDKVGTGAFAFPIEEYVHVVEEHFOLLARRMOVDKRVYLATDDPTLLKE 416
DB 121 HPVIGVHVARTDKVGTGAFAFPIEEYVHVVEEHFOLLARRMOVDKRVYLATDDPTLLKE 180
QY 417 ATKYSNYEPISDNISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAY 476
DB 181 ATKYPNYEPISDNISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAY 240
QY 477 EIMQTLHPDASANFHSLLDDIYVFGQNAHNOIAVYPHKPRTEETPMPEPGDIIIGVAGNHW 536
DB 241 EIMQTLHPDASANFHSLLDDIYVFGQNAHNOIAVYPHKPRTEETPMPEPGDIIIGVAGNHW 300
QY 537 DGYSKGINRKLGTGLPSYKVKREKTIETVKYPTYPEAEK 575
DB 301 DGYSKGVNRKLGTGLPSYKVKREKTIETVKYPTYPEAEK 339
RESULT 8
ID ABB59117 standard; Protein; 619 AA.
XX
XX ABB59117;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 4143.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD

```

```

XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL03220.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 4143; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 619 AA;
SQ
Query Match 48.0%; Score 1480.5; DB 22; Length 619;
Best Local Similarity 46.5%; Pred. No. 4e-129;
Matches 285; Conservative 108; Mismatches 169; Indels 51; Gaps 8;
QY 3 AWTGSM-RWIMLILFANGTLLFYIGHLVRDN-----DHPDHSRELSKILAKLERLKKQ 56
DB 10 ASANSWARALITIFLWAGLIVGVFVVKLTNTQGOAAGESELNARRISQALQMLEHTQR 69
QY 57 NEDLRRAESLRIPGPIDQGTATGRVRLVEQLVKAKEIE----- 98
DB 70 NEELKQLIDELMSDQ--LDKQSAKMLVORLENDALNPKLAPEVAGPESFESAPADLR 127
QY 99 ---NYKKQARNGL-----GKDHEILRRRIENGAKELWFFLQSELSKCLK----- 138
DB 128 GWNVVAEGAPNDLEAGVPDHCHEFEPSELEYETFRRIQTNGEINWFFSSELGKVRKAVAA 187
QY 139 -HLEGNELQRHADIILLDLGHHERSIMTDLYLLYSQTDGAGDWREKAOKDLTELVRRTY 197
DB 188 GHASA-DLEESINQVLLQGAHEKRSLLSDMERMQSDGYEAWRHKEARDLSDLVQRRLHH 246
QY 198 LQNPDCSKARKLYCNINKGCGYCOLHHVYVCEWIAVGTORTLILSQNRWYATGGWET 257
DB 247 LQNPEDCONARKLVCKLNKCGYCOLHHVYVCFIVAYATERILLRSGRWYRHKGWEE 306
QY 258 VFRPVSETCTDRSGLSTGHWSGEVNDKNIQVVELPIVDSLHPRPPLVLPVDELADRL 317
DB 307 VFPVSNVSCHDAGTANTYNNWPKG---PNTQVLVLPIDSLMPPRPPYLPVDELAPL 363
QY 318 RVHGDPAVWVVSQFVKYLIRPQWLEKEIEEATKLGFKHPVIGVHVARTDKVGTGAFAFH 377
DB 364 RLHGDPVWVVGQFLKYLRLRPQPTTRDFTLSTGMENLWGERPIVGHVHVRTDKVGTGAACH 423
QY 378 PIEEYVHVVEEHFOLLARRMOVDKRVYLATDDPTLLKEAKTKYSNYEPISDNISWSAG 437
DB 424 SVEEYMTYVEDYRTLEVNGSTVARRIFLASDDDAQVIEEARKKTPQGIIGDPEVARNAS 483
QY 438 LHNRYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFHSLLDDIY 497
DB 484 VSTRYTDALNGIILDIHLLSMSDLHVLCTFSSQVCRVAYEIMQTMYPDAHRFKSLDDIY 543

```



```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPVIG 361
Db 61 PWLEKEIEEATKKLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPVIG 361
Db 61 PWLEKEIEEATKKLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPVIG 361
Db 61 PWLEKEIEEATKKLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPVIG 361
Db 61 PWLEKEIEEATKKLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```

```
ABE28797
ID ABB28797 standard; Peptide; 82 AA.
XX ABB28797;
AC
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00662.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 11765; 327pp + sequence listing; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 82 AA;
SQ
Query Match 14.0%; Score 432; DB 22; Length 82;
Best Local Similarity 96.3%; Pred. No. 2e-32;
Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EVNDKNIQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPVIG 361
Db 61 PWLEKEIEEATKKLGFKHPVIG 82
RESULT 13
ABBI9422
ID ABB19422 standard; Protein; 82 AA.
XX
XX ABB19422;
XX
XX 23-JAN-2002 (first entry)
DT
```





```
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 27445; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX SQ Sequence 82 AA;
XX Query Match 14.0%; Score 432; DB 22; Length 82;
XX Best Local Similarity 96.3%; Pred. No. 2e-32;
XX Matches 79; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 280 EYNDKNIQVVELPIVDSLHPRPPYLPVLPEDLADRLRVHGDPAVWVVSQFVKYLIRPQ 339
Db 1 EVKDKNVQVVELPIVDSLHPRPPYLPVLPEDLADRLRVHGDPAVWVVSQFVKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFQHPVIG 361
Db 61 PWLEKEIEEATKKLGFQHPVIG 82
```

Search completed: February 2, 2004, 08:41:47  
Job time : 39.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 15.5 Seconds  
(without alignments)  
1569.596 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRAWTGSWRWIMLILFAWT.....YKREKIETVKYTPPEAK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3003	97.3	575	3	US-08-913-805A-10
2	3003	97.3	575	3	US-09-442-629-10
3	2910	94.3	575	3	US-08-913-805A-2
4	2910	94.3	575	3	US-09-442-629-2
5	131	4.2	25	3	US-08-913-805A-12
6	131	4.2	25	3	US-09-442-629-12
7	114	3.7	214	5	PCT-US96-08950-2
8	114	3.7	214	5	PCT-US96-09127-2
9	114	3.7	229	3	US-08-630-915A-221
10	113.5	3.7	976	3	US-09-104-324B-4
11	113	3.7	26	3	US-08-913-805A-3
12	113	3.7	26	3	US-09-442-629-3
13	113	3.7	3248	1	US-08-353-700-1
14	113	3.7	3248	5	PCT-US95-16216-1
15	110.5	3.6	644	4	US-09-198-452A-63
16	109	3.5	352	4	US-08-630-915A-12
17	109	3.5	1230	2	US-08-968-542C-35
18	107	3.5	418	4	US-09-198-452A-1056
19	104	3.4	251	4	US-08-630-915A-8
20	104	3.4	677	3	US-08-836-567-2
21	104	3.4	677	3	US-09-606-304-2
22	104	3.4	1197	3	US-08-836-567-12
23	104	3.4	1197	4	US-09-606-304-12
24	103	3.3	575	4	US-09-107-532A-3910
25	102.5	3.3	2482	1	US-08-328-254-6
26	101	3.3	1027	4	US-09-914-259-27
27	101	3.3	1939	4	US-09-310-187A-1

Sequence 2, Appli  
Sequence 2, Appli  
Sequence 3877, Ap  
Sequence 11, Appli  
Sequence 2, Appli  
Sequence 67, Appl  
Sequence 3, Appli  
Sequence 89, Appl  
Sequence 4794, Ap  
Sequence 2, Appli  
Sequence 11, Appl  
Sequence 16, Appl  
Sequence 35, Appl  
Sequence 3, Appli  
Sequence 12, Appl  
Sequence 67, Appl  
Sequence 2, Appli  
Sequence 75, Appl

28 101 3.3 2185 2 US-08-822-445-2  
29 101 3.3 2185 4 US-09-396-540-2  
30 100 3.2 381 4 US-09-107-532A-3877  
31 100 3.2 3878 4 US-09-914-259-11  
32 99.5 3.2 640 4 US-08-873-404-2  
33 99.5 3.2 959 4 US-09-914-259-67  
34 99 3.2 366 4 US-09-750-580-3  
35 99 3.2 366 4 US-09-599-360B-89  
36 98 3.2 1037 4 US-09-134-001C-4794  
37 98 3.2 1005 2 US-08-935-450-2  
38 98 3.2 1588 5 PCT-US93-07261-11  
39 98 3.2 1663 5 PCT-US93-07261-16  
40 97.5 3.2 353 1 US-07-752-101A-35  
41 97.5 3.2 466 4 US-09-610-401-3  
42 97.5 3.2 466 4 US-09-167-206-12  
43 97.5 3.2 1646 4 US-09-535-008-67  
44 97.5 3.2 1647 4 US-09-535-008-2  
45 97.5 3.2 1649 4 US-09-535-008-75

## ALIGNMENTS

RESULT 1  
US-08-913-805A-10  
; Sequence 10, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Neofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

```
; MOLECULE TYPE: peptide
US-08-913-805A-10

Query Match
Best Local Similarity 97.3%; Score 3003; DB 3; Length 575;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLIFYGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLIFYGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60

QY 61 RMAESLRIPGPIDQGTATGRVRLVEEQVAKAEQIENYKQARNGLGKDHILRRRIE 120
DB 61 RMAESLRIPGPIDQGTATGRVRLVEEQVAKAEQIENYKQARNGLGKDHILRRRIE 120

QY 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180
DB 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180

QY 181 EKEAKDLTELVRITYLQNPDKCSKARKLVNKNKGCGYGCQLHHVVCYCFMIAVGTQRT 240
DB 181 EKEAKDLTELVRITYLQNPDKCSKARKLVNKNKGCGYGCQLHHVVCYCFMIAVGTQRT 240

QY 241 LILESQNRVATGWTVPVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRVATGWTVPVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVHGDPAVWWSQVVKYLIRPQWLEKEIEATKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLRVHGDPAVWWSQVVKYLIRPQWLEKEIEATKLGFKHPVI 360

QY 361 GVHVRRDKVGTAAAFHPIEEYVWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420
DB 361 GVHVRRDKVGTAAAFHPIEEYVWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420

QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480

QY 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575
DB 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575

RESULT 2
US-09-442-629-10
; Sequence 10, Application US/09442629
; Patent No. 6291219
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; UOZUMI, Naofumi
; SHIBA, Tetsuo
; YANAGIDANI, Shusaku
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,629
```

```
; FILING DATE: 18-No. 6291219-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/913.805A
; FILING DATE: 7 JAN 1998
; APPLICATION NUMBER: PCT/JP97/00171
; FILING DATE: 23 JAN 1997
; APPLICATION NUMBER: JP 192260
; FILING DATE: 22 JUL 1996
; APPLICATION NUMBER: JP 162813
; FILING DATE: 24 JUN 1996
; APPLICATION NUMBER: JP 161648
; FILING DATE: 21 JUN 1996
; APPLICATION NUMBER: JP 10365
; FILING DATE: 24 JAN 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2356/3
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-442-629-10

Query Match 97.3%; Score 3003; DB 3; Length 575;
Best Local Similarity 96.5%; Pred. No. 4.5e-274;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRWTGSRWIMLILFAWGTLIFYGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60
DB 1 MRPTGSRWIMLILFAWGTLIFYGHLVRDNDHPDHSRSLSKILAKLERLKQONEDL 60

QY 61 RMAESLRIPGPIDQGTATGRVRLVEEQVAKAEQIENYKQARNGLGKDHILRRRIE 120
DB 61 RMAESLRIPGPIDQGTATGRVRLVEEQVAKAEQIENYKQARNGLGKDHILRRRIE 120

QY 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180
DB 121 NGAKELWFFLQSELKKLHLEGNELQRADEFLDLGHHERSIMTDLVYLSQTDGAGDWR 180

QY 181 EKEAKDLTELVRITYLQNPDKCSKARKLVNKNKGCGYGCQLHHVVCYCFMIAVGTQRT 240
DB 181 EKEAKDLTELVRITYLQNPDKCSKARKLVNKNKGCGYGCQLHHVVCYCFMIAVGTQRT 240

QY 241 LILESQNRVATGWTVPVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRVATGWTVPVSETCTDRSGISTGHWSGEVNDKNIQVVELPIVDSLHPR 300

QY 301 PPYLPLAVPEDLADRLRVHGDPAVWWSQVVKYLIRPQWLEKEIEATKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLRVHGDPAVWWSQVVKYLIRPQWLEKEIEATKLGFKHPVI 360

QY 361 GVHVRRDKVGTAAAFHPIEEYVWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420
DB 361 GVHVRRDKVGTAAAFHPIEEYVWVVEHFQLLARRMQVKKRVYLATDDPTLLKEATK 420

QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480

QY 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575
DB 541 KGINRKLGTGLYPSYKVKREKIEVTKYPTYPEAEK 575
```

## RESULT 3

US-08-913-805A-2  
; Sequence 2, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913,805A  
; FILING DATE: 7 JAN 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP97/00171  
; FILING DATE: 23 JAN 1997  
; APPLICATION NUMBER: JP 192260  
; FILING DATE: 22 JUL 1996  
; APPLICATION NUMBER: JP 162813  
; FILING DATE: 24 JUN 1996  
; APPLICATION NUMBER: JP 161648  
; FILING DATE: 21 JUN 1996  
; APPLICATION NUMBER: JP 10365  
; FILING DATE: 24 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2356/3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 575 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-913-805A-2

Query Match 94.3%; Score 2910; DB 3; Length 575;

Best Local Similarity 93.6%; Pred. No. 2.6e-265;

Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY	1	MRATGSRWRIMLILFANGTLLFYIGGHLVRDNDHSDHSSRELKILAKLERLKQONEDL	60
DB	1	MRPTGSRWRIMLILFANGTLLFYIGGHLVRDNDHSDHSSRELKILAKLERLKQONEDL	60
QY	61	RMASLSRIPEGPIDQGTATGVRVLEQLVAKQIQENYKQANGLGKQKHILRRRIE	120
DB	61	RMASLSRIPEGPIDQGPASGRVRALEQFMKAEQIQENYKQTKNGPKQKHILRRRIE	120
QY	121	NGAKELWFLQSELKGLHLEGNLORHADETLILLGHHSRIMTDLYVLSQTDGAGDWR	180
DB	121	NGAKELWFLQSELKGLHLEGNLORHADEFLSLGHHSRIMTDLYVLSQTDGAGDWR	180
QY	181	EKEAKDLTELVRRTTYLQNPDKCSKAKKLVNKNKGGCGYCOLHHVVVCFMAYGTORT	240

Db	181	EKEAKDLTELVRRTTYLQNPDKCSKAKKLVNKNKGGCGYCOLHHVVVCFMAYGTORT	240
QY	241	LILESQNRVATGNETVFRPVSETCTDRSGLSGTHWSGEVNDKNIQVVELPIVDSLHPR	300
Db	241	LALSHNRVATGNETVFRPVSETCTDRSGLSGTHWSGEVNDKNIQVVELPIVDSLHPR	300
QY	301	PPYLPVAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBATKGLGKHPVI	360
Db	301	PPYLPVAVPEDLADRLVRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBATKGLGKHPVI	360
QY	361	GVHVRDVKVCTEAAPHIEEYMHVHEHFOLLARRMQVDKRVYLATDDPTLLKEATK	420
Db	361	GVHVRDVKVGAFAFPIEETVHVEEDFOLLARRMQVDKRVYLATDDPTLLKEATK	420
QY	421	YSNYEFISDSISWSAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ	480
Db	421	YPSYEFISDSISWSAGLHNYTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ	480
QY	481	TLHPDASANFSLDDIYVFGGQNAHQIAVYPHPRTEEEIPMBPGDIIGVAGNHWGYS	540
Db	481	ALHPDASANFSLDDIYVFGGQNAHQIAVYPHPRTEEEIPMBPGDIIGVAGNHWGYP	540
QY	541	KGINRKLKGLTGLYSYKVKREKIEIVKYPTYPEAEK	575
Db	541	KGVNRKLKGLTGLYSYKVKREKIEIVKYPTYPEADK	575

## RESULT 4

US-09-442-629-2  
; Sequence 2, Application US/09442629  
; Patent No. 6291219

## GENERAL INFORMATION:

APPLICANT: TANIGUCHI, Naoyuki

UOZUMI, Naofumi

SHIBA, Tetsuo

YANAGIDANI, Shusaku

TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

CITY: Washington

STATE: DC

COUNTRY: US

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2

SOFTWARE: WordPerfect 6.1 Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/442,629

FILING DATE: 18-NO. 6291219-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/913,805A

FILING DATE: 7 JAN 1998

APPLICATION NUMBER: PCT/JP97/00171

FILING DATE: 23 JAN 1997

APPLICATION NUMBER: JP 192260

FILING DATE: 22 JUL 1996

APPLICATION NUMBER: JP 162813

FILING DATE: 24 JUN 1996

APPLICATION NUMBER: JP 161648

FILING DATE: 21 JUN 1996

APPLICATION NUMBER: JP 10365

FILING DATE: 24 JAN 1996

ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.

REGISTRATION NUMBER: 39,048

REFERENCE/DOCKET NUMBER: 2356/3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-429-1776

TELEFAX: 202-429-0796

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-442-629-2

Query Match 94.3%; Score 2910; DB 3; Length 575;  
Best Local Similarity 93.6%; Pred. No. 2.6e-265;  
Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MRWGTGSRWIMLLIFAWGTLFYIGGHLVRNDHPDHSRSLKILAKLERLKQONEDL 60  
DB 1 MRWGTGSRWIMLLIFAWGTLFYIGGHLVRNDHSRSLKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPGIDOGTATGRVRLVEQLVKAKQIENYKQARNGLGKQHEILRRRIE 120  
DB 61 RMAESLRIPGIDOGTATGRVRLVEQLVKAKQIENYKQARNGLGKQHEILRRRIE 120  
QY 121 NGAKELWFFLQSELKHLKLEGNELQORHADELTLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
DB 121 NGAKELWFFLQSELKHLKLEGNELQORHADELTLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
QY 181 EKEAKDLTELVRRTYTLQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMAYGTQRT 240  
DB 181 EKEAKDLTELVRRTYTLQNPDKCSKARKLVNINKGCGYGCQLHHVYVCFMAYGTQRT 240  
QY 241 LILESQWRYATGWTETPRPVSETCTDRSGI-STGHSGEVNDKNIQVVELPVDVSLHPR 300  
DB 241 LALESHNRYATGWTETPRPVSETCTDRSGI-STGHSGEVNDKNIQVVELPVDVSLHPR 300  
QY 301 PPVPLAVPEDLADRLRLVHGPVAVWWSQFVKYLIRPQWLEKEIEEATKLGFKHPVI 360  
DB 301 PPVPLAVPEDLADRLRLVHGPVAVWWSQFVKYLIRPQWLEKEIEEATKLGFKHPVI 360  
QY 361 GVHVRTDKVGTAAFPHTPEEYVHVVEHFQLLARRMQVDKRVYLAITDDPTLLKEATK 420  
DB 361 GVHVRTDKVGTAAFPHTPEEYVHVVEHFQLLARRMQVDKRVYLAITDDPTLLKEATK 420  
QY 421 YSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
DB 421 YSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFSLDDIYFEGGNAHNOIAVYHPKRTTEEIPWEPGDIIGVAGNHWDCYS 540  
DB 481 ALHPDASANFRSLDDIYFEGGNAHNOIAVYHPKRTTEEIPWEPGDIIGVAGNHWDCYS 540  
QY 541 KGINRKLGTGLYPSYKVKREKIEYKPYTPAEK 575  
DB 541 KGVNRLKGTGLYPSYKVKREKIEYKPYTPAEK 575

RESULT 5  
US-08-913-805A-12  
; Sequence 12, Application US/08913805A  
; Patent No. 6054304  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913.805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-12

Query Match 4.2%; Score 131; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 352 KLGFKHPVIGVHVVRTDKVGTAAAF 376  
DB 1 KLGFKHPVIGVHVVRTDKVGTAAAF 25

RESULT 6  
US-09-442-629-12  
; Sequence 12, Application US/09442629  
; Patent No. 6291219  
; GENERAL INFORMATION:  
; APPLICANT: TANIGUCHI, Naoyuki  
; APPLICANT: UOZUMI, Naofumi  
; APPLICANT: SHIBA, Tetsuo  
; APPLICANT: YANAGIDANI, Shusaku  
; TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/442.629  
; FILING DATE: 18-No. 6291219-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/913.805A  
; FILING DATE: 7 JAN 1998  
; APPLICATION NUMBER: PCT/JP97/00171

;; FILING DATE: 23 JAN 1997  
;; APPLICATION NUMBER: JP 192260  
;; FILING DATE: 22 JUL 1996  
;; APPLICATION NUMBER: JP 162813  
;; FILING DATE: 24 JUN 1996  
;; APPLICATION NUMBER: JP 161648  
;; FILING DATE: 21 JUN 1996  
;; APPLICATION NUMBER: JP 10365  
;; FILING DATE: 24 JAN 1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Toffenetti, Judith L.  
;; REGISTRATION NUMBER: 39,048  
;; REFERENCE/DOCKET NUMBER: 2356/3  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-429-1776  
;; TELEFAX: 202-429-0796  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-442-629-12

Query Match 4.2%; Score 131; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 KLGFKHPVIGVHVRTDKVGTEAAF 376  
Db 1 KLGFKHPVIGVHVRTDKVGTEAAF 25

RESULT 7  
PCT-US96-08950-2  
; Sequence 2, Application PC/TUS9608950  
; GENERAL INFORMATION:  
; APPLICANT: OSTEOSA INC.  
; TITLE OF INVENTION: Osteoclast Growth Regulatory Factor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/08950  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul C. Steinhart  
; REGISTRATION NUMBER: 30,806  
; REFERENCE/DOCKET NUMBER: FP-ND 2121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US96-08950-2

Query Match 3.7%; Score 114; DB 5; Length 214;

Best Local Similarity 38.0%; Pred. No. 0.012;  
Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;  
Qy 509 AVYPHKPRTTEEIPMEPGDIIGVAG-----NHWGYSKGINKLKGKTGLYPSYKVREK1ET 564  
Db 19 ALYTFEPTDELYFEEGDIIYITDMSDTNWKTSK-----GRTGLIPSNYYVAEQAES 72  
Qy 565 VKYPTYPEAK 575  
Db 73 IDNPLHEAAKR 83

RESULT 8  
PCT-US96-09127-2  
; Sequence 2, Application PC/TUS9609127  
; GENERAL INFORMATION:  
; APPLICANT: OSTEOSA INC.  
; TITLE OF INVENTION: Osteoclast Growth Regulatory Factor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/09127  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul C. Steinhart  
; REGISTRATION NUMBER: 30,806  
; REFERENCE/DOCKET NUMBER: FP-ND 2122  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 214 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US96-09127-2

Query Match 3.7%; Score 114; DB 5; Length 214;  
Best Local Similarity 38.0%; Pred. No. 0.012;  
Matches 27; Conservative 12; Mismatches 22; Indels 10; Gaps 2;  
Qy 509 AVYPHKPRTTEEIPMEPGDIIGVAG-----NHWGYSKGINKLKGKTGLYPSYKVREK1ET 564  
Db 19 ALYTFEPTDELYFEEGDIIYITDMSDTNWKTSK-----GRTGLIPSNYYVAEQAES 72  
Qy 565 VKYPTYPEAK 575  
Db 73 IDNPLHEAAKR 83

RESULT 9  
US-08-630-915A-221  
; Sequence 221, Application US/08630915A  
; Patent No. 6309820  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: HOFFMAN, No. 6309820h  
; APPLICANT: KAY, Brian K.  
; APPLICANT: FOWLKES, Dana M.  
; APPLICANT: MCCONNELL, Stephen J.





ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A  
FILING DATE: 7 JAN 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-805A-3

Query Match 3.7%; Score 113; DB 3; Length 26;  
Best Local Similarity 88.0%; Pred. No. 0.00054;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 102 KOARNGLGKDHILRRRIENGAKEL 126  
DB 1 KQTKNGPGKHILRRRIENGAKEL 25  
||:||||||||||||||||||  
||:||||||||||||||||||  
  
RESULT 12  
US-09-442-629-3  
Sequence 3, Application US/09442629  
Patent No. 6291219  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
UOZUMI, Naofumi  
SHIBA, Tetsuo  
YANAGIDANI, Shusaku  
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/442,629  
FILING DATE: 18-NO. 6291219-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,805A

FILING DATE: 7 JAN 1998  
APPLICATION NUMBER: PCT/JP97/00171  
FILING DATE: 23 JAN 1997  
APPLICATION NUMBER: JP 192260  
FILING DATE: 22 JUL 1996  
APPLICATION NUMBER: JP 162813  
FILING DATE: 24 JUN 1996  
APPLICATION NUMBER: JP 161648  
FILING DATE: 21 JUN 1996  
APPLICATION NUMBER: JP 10365  
FILING DATE: 24 JAN 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2356/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-442-629-3  
  
Query Match 3.7%; Score 113; DB 3; Length 26;  
Best Local Similarity 88.0%; Pred. No. 0.00054;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 102 KOARNGLGKDHILRRRIENGAKEL 126  
DB 1 KQTKNGPGKHILRRRIENGAKEL 25  
||:||||||||||||||||||  
||:||||||||||||||||||  
  
RESULT 13  
US-08-353-700-1  
Sequence 1, Application US/08353700  
Patent No. 5599919  
GENERAL INFORMATION:  
APPLICANT: YEN, TIMOTHY J.  
TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
TITLE OF INVENTION: AND METHODS OF USE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
STREET: 1601 MARKET STREET, SUITE 720  
CITY: PHILADELPHIA  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
;
US-08-353-700-1

Query Match          3.7%; Score 113; DB 1; Length 3248;
Best Local Similarity 19.3%; Pred. No. 1;
Matches 102; Conservative 86; Mismatches 176; Indels 164; Gaps 26;

QY 29 LVRDNDHPDHSRSLKILAKLERLKKQONEDL---RRMAESLRIPGIDQGTATGRVRV 85
DB 343 LVRTTAQYDQAS---TKYTALEQKLKLTEDLSQORQNAESARCS-----384

QY 86 LEEQL-VKAEQIENYKKQARNGLKGDHEILRRRIE-----NGAKELWFFLOSELKCLKH 139
DB 385 LEQIKKEKEFEQELSQRQSFOTLQECIQMKARLTQELQQAQNMHNVLQAEILDKLT 444

QY 140 LEGNELQRHADIILLDLGHHRSIMTDLVYLSQTDGADWREKAQDLTELVRITVQL 199
DB 445 VK-QOLENNLEEFKQKLCRAEAFQ-----ASQIKENELRRSMEEMKKNLLK 492

QY 200 NPKDCSKARKLVNINCGCGYCOLHHVYCFMAYGTQRTLILESQNW-----RYATGGW 255
DB 493 SHSE-QKARE-VCHLE-----AELKNIKQC-----LNQSQNFABEMKAKNTSQ 533

QY 256 ETVPFVSECTDRSGLSTGHSWGEVNDKNIQVVELPIV-----DSLHPRPPVLP 305
DB 534 ETMLRDLQEKINQOE-----NSLTLEKLVADLEKQDCSQDLKKREHHI- 581

QY 306 LAVPEDLADRLLRVHGDPVAV-----WVVSQFVKYLIRPQWL 342
DB 582 ----EQNDKLSKTEKESKALLSLELKKKEYEELKEEKTLPSCWKSENEKLLTQ----M 633

QY 343 EKEIEEATKKGFGHPVIGVHVRRTDKVGTAAFPPIEYVMHVE---EHFOLLARRMQ- 398
DB 634 ESEKENLQSKINHLETCL-----KTQQIKS-----HEYNERVRLTMDRENLSVEIRNLHN 684

QY 399 -VDKRVVLATDDPTLLK-EAKTKSYNEFFSDNSISWSAGLHNRVTEN-----SLRGV 450
DB 685 VLDSKSVEVETQKLAYMELQKAEFSQK-----HKEIENMCLTKTSQLTGQ 731

QY 451 ILDIHFLSQADFLVCTFSSQVCRVAYEIM--QTLHPDASANFHSLLDI 496
DB 732 VEDLEHKLQL-----LSNEIMDKRCYQDLHAEYESLRDL 766
```

```
RESULT 14
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
```

```
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match          3.7%; Score 113; DB 5; Length 3248;
Best Local Similarity 19.3%; Pred. No. 1;
Matches 102; Conservative 86; Mismatches 176; Indels 164; Gaps 26;

QY 29 LVRDNDHPDHSRSLKILAKLERLKKQONEDL---RRMAESLRIPGIDQGTATGRVRV 85
DB 343 LVRTTAQYDQAS---TKYTALEQKLKLTEDLSQORQNAESARCS-----384

QY 86 LEEQL-VKAEQIENYKKQARNGLKGDHEILRRRIE-----NGAKELWFFLOSELKCLKH 139
DB 385 LEQIKKEKEFEQELSQRQSFOTLQECIQMKARLTQELQQAQNMHNVLQAEILDKLT 444

QY 140 LEGNELQRHADIILLDLGHHRSIMTDLVYLSQTDGADWREKAQDLTELVRITVQL 199
DB 445 VK-QOLENNLEEFKQKLCRAEAFQ-----ASQIKENELRRSMEEMKKNLLK 492

QY 200 NPKDCSKARKLVNINCGCGYCOLHHVYCFMAYGTQRTLILESQNW-----RYATGGW 255
DB 493 SHSE-QKARE-VCHLE-----AELKNIKQC-----LNQSQNFABEMKAKNTSQ 533

QY 256 ETVPFVSECTDRSGLSTGHSWGEVNDKNIQVVELPIV-----DSLHPRPPVLP 305
DB 534 ETMLRDLQEKINQOE-----NSLTLEKLVADLEKQDCSQDLKKREHHI- 581

QY 306 LAVPEDLADRLLRVHGDPVAV-----WVVSQFVKYLIRPQWL 342
DB 582 ----EQNDKLSKTEKESKALLSLELKKKEYEELKEEKTLPSCWKSENEKLLTQ----M 633

QY 343 EKEIEEATKKGFGHPVIGVHVRRTDKVGTAAFPPIEYVMHVE---EHFOLLARRMQ- 398
DB 634 ESEKENLQSKINHLETCL-----KTQQIKS-----HEYNERVRLTMDRENLSVEIRNLHN 684

QY 399 -VDKRVVLATDDPTLLK-EAKTKSYNEFFSDNSISWSAGLHNRVTEN-----SLRGV 450
DB 685 VLDSKSVEVETQKLAYMELQKAEFSQK-----HKEIENMCLTKTSQLTGQ 731

QY 451 ILDIHFLSQADFLVCTFSSQVCRVAYEIM--QTLHPDASANFHSLLDI 496
DB 732 VEDLEHKLQL-----LSNEIMDKRCYQDLHAEYESLRDL 766
```

```
RESULT 15
US-09-198-452A-63
; Sequence 63, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
```

;; CURRENT APPLICATION NUMBER: US/09/198.452A  
;; CURRENT FILING DATE: 1998-11-24  
;; NUMBER OF SEQ ID NOS: 6849  
;; SEQ ID NO 63  
;; LENGTH: 644  
;; TYPE: PRT  
;; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-63

Query Match 3.6%; Score 110.5; DB 4; Length 644;  
Best Local Similarity 21.0%; Pred. No. 0.14; Indels 125; Gaps 20;  
Matches 103; Conservative 72; Mismatches 191; Indels 125; Gaps 20;  
Qy 43 LSKILAKLERLKQNMEDLRMAESIRIPEGPIDQCTA-----TGRVRVLEQLVKAKEQI 97  
Db 75 ISGFLLLLERREVSQVLEGIPGTG--IPVPSAEPSSSEIQKKQKAKQILDLQELDQL 132  
Qy 98 ENYKKQARNGLG--KDHEILRRRIENGAK-----LWFFLOSELKKL-----KHL 140  
Db 133 DTDIQHVLSCICLGLKDLKCKDRGLLKDAKEKIQVDFVWKDMMMEFVELQQVMDQESRYL 192  
Qy 141 EG--NELQRHADEILLDLGHHERSIMTDLYYLSQTDGAGDWREKEAKDLTELQVRRITYL 198  
Db 193 EGLIHEVQSIHAKLFVDDVNTIRSHLGESCGYLPSEDVRGELLKRFAP---EVAR---FM 246  
Qy 199 QNPDKCSKARKLVNCNKGCGYQGLHHVYCFMIAYGTQRTLILESQNMRYA----- 251  
Db 247 KVTED---IRKIAPFNKN--AYGAKN---AFDKAFGSLCTCLYKSLTKSYRDTFCDYK 298  
Qy 252 -----TGGWETVPRPVSETCTDRSGLSTGHWSGEVNDKNIQVVLPFIVDSLHPR- 300  
Db 299 RAKILPDENNARSARAEQRFREVKDHWEDLN--ETVFWVKE--DGRIDIEVLTAVGWPDY 354  
Qy 301 PPYLPPLAVPED-----LADRLLRVHGDPVWVWSQFVKYLIRPQWLEKEIEATK--- 351  
Db 355 PEHLILEKRRKDKVMSHQLWEATMRVKEAEVTSVAR-VAFEKDGSGQQNKKFQEKTKERL 413  
Qy 352 -----KLGFHPVIGVHVRTD---KVGTEAAPHPPIEE----- 381  
Db 414 RCLKDLRDOECHRAQERLEKLTALYPEVSVSVVETERKFNLEKAYGNLEERYQSVVQD 473  
Qy 382 -----YMHVVEEHFQLLARMQVDRKRVYLATDDPTLLKEAK 418  
Db 474 QEDYWTEQKNREAEFRAGTKVRSMEEEVAEHLQILENLEDCYKRLSKAETFALGVVERA 533  
Qy 419 TKYSNYEFISD 429  
Db 534 TEEIEYTLISD 544

Search completed: February 2, 2004, 08:44:37  
Job time : 16.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 2, 2004, 08:40:23 ; Search time 217 Seconds  
(without alignments)  
550.686 Million cell updates/sec

Title: US-09-971-773-24

Perfect score: 3085

Sequence: 1 MRAWTGSWRIMLILFWANGT.....YKVRKTIETVKYPTYPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3085	100.0	575	11	US-09-971-773-24
2	3024	98.0	575	11	US-09-971-773-23
3	2992	97.0	575	9	US-09-839-136-10
4	2903	94.1	575	9	US-09-839-136-2
5	2658	86.2	515	15	US-10-106-698-4658
6	432	14.0	82	9	US-09-864-761-34720
7	426	13.8	81	9	US-09-864-761-46107
8	127	4.1	1055	12	US-10-369-493-12504
9	123	4.0	485	12	US-10-104-047-3419
10	122	4.0	25	9	US-09-839-136-12
11	118.5	3.8	976	12	US-10-117-937-596
12	116.5	3.8	919	12	US-10-369-493-22825
13	116.5	3.8	1992	12	US-10-369-493-6527
14	116	3.8	678	12	US-10-369-493-5427
15	116	3.8	1285	15	US-10-205-823-273

16	116	3.8	1285	15	US-10-177-293-317	Sequence 317, App
17	115	3.7	1137	12	US-10-369-493-6931	Sequence 6931, Ap
18	115	3.7	1156	12	US-10-369-493-43	Sequence 43, Appl
19	114.5	3.7	898	11	US-09-893-519A-73	Sequence 73, Appl
20	114	3.7	223	9	US-09-879-957-221	Sequence 221, App
21	113.5	3.7	2020	12	US-10-369-493-5128	Sequence 5128, Ap
22	113.5	3.7	2020	12	US-10-369-493-5129	Sequence 5129, Ap
23	113	3.7	26	9	US-09-839-136-3	Sequence 3, Appli
24	113	3.7	405	12	US-10-094-749-3191	Sequence 3191, Ap
25	113	3.7	1959	12	US-10-028-248A-36	Sequence 36, Appl
26	111.5	3.6	1961	12	US-10-028-248A-105	Sequence 105, App
27	111	3.6	1441	12	US-10-412-897-3	Sequence 3, Appli
28	111	3.6	1597	14	US-10-017-216-6	Sequence 6, Appli
29	111	3.6	1641	14	US-10-017-216-5	Sequence 5, Appli
30	111	3.6	2055	14	US-10-017-216-4	Sequence 4, Appli
31	110.5	3.6	644	12	US-10-289-762-63	Sequence 63, Appl
32	110.5	3.6	1979	15	US-10-205-823-419	Sequence 419, App
33	110	3.6	586	12	US-10-097-111-286	Sequence 286, App
34	109.5	3.5	262	12	US-10-032-189-79	Sequence 79, Appl
35	109.5	3.5	1179	12	US-10-369-493-1330	Sequence 1330, Ap
36	109.5	3.5	1179	12	US-10-369-493-20394	Sequence 20394, A
37	109.5	3.5	1978	12	US-10-094-466-64	Sequence 64, Appl
38	109	3.5	352	9	US-09-879-957-12	Sequence 12, Appl
39	109	3.5	1001	12	US-10-262-445-66	Sequence 66, Appl
40	109	3.5	1068	12	US-10-262-445-64	Sequence 64, Appl
41	108.5	3.5	521	12	US-10-369-493-5956	Sequence 5956, Ap
42	108	3.5	880	12	US-10-369-493-21643	Sequence 21643, A
43	107	3.5	418	12	US-10-289-762-1056	Sequence 1056, Ap
44	107	3.5	863	12	US-10-369-493-10297	Sequence 10297, A
45	106	3.4	807	12	US-10-132-350-42	Sequence 42, Appl

## ALIGNMENTS

## RESULT 1

US-09-971-773-24  
; Sequence 24, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuyasu UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: NO. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-971-773-24

Query Match 100.0%; Score 3085; DB 11; Length 575;  
Best Local Similarity 100.0%; Pred. No. 2.3e-271;  
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAWTGSWRIMLILFWANGTLLFYGGHLYRDNDHPHSSRELSKIHLKLRKQOEDL 60  
Db 1 MRAWTGSWRIMLILFWANGTLLFYGGHLYRDNDHPHSSRELSKIHLKLRKQOEDL 60  
Qy 61 RRMAESLRIPGPDQGTATGRVRLVEQLVKAKEQIENYKQARNGIKGDKHILRRRIE 120

Db 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHEILRRRIE 120  
Qy 121 NGAKELWFFLQSELKKLHLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Db 121 NGAKELWFFLQSELKKLHLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Qy 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCNKGCGYGCQLHHVYVCFMIAVGTQRT 240  
Db 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCNKGCGYGCQLHHVYVCFMIAVGTQRT 240  
Qy 241 LILESQNRWYATGCHWETVFRPVSCTDRSGLSGTHSGSEVNDKNIQVWELPIVDSLHPR 300  
Db 241 LILESQNRWYATGCHWETVFRPVSCTDRSGLSGTHSGSEVNDKNIQVWELPIVDSLHPR 300  
Qy 301 PYPPLAVPEDLADRLRLRHVHGDPAVMWVSQFVKYLIRPQPMLEKIEBEATKLGFKHPVI 360  
Db 301 PYPPLAVPEDLADRLRLRHVHGDPAVMWVSQFVKYLIRPQPMLEKIEBEATKLGFKHPVI 360  
Qy 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEAKTK 420  
Db 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEAKTK 420  
Qy 421 YSNYEFISDNTSWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YSNYEFISDNTSWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFHLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Db 481 TLHPDASANFHLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Qy 541 KGINRKLGTGLPSYKVKREKIEVTKYPTYPEAEK 575  
Db 541 KGINRKLGTGLPSYKVKREKIEVTKYPTYPEAEK 575

## RESULT 2

US-09-971-773-23  
; Sequence 23, Application US/09971773  
; Publication No. US20030115614A1  
; GENERAL INFORMATION:  
; APPLICANT: Yutaka KANDA  
; APPLICANT: Mitsuo SATOH  
; APPLICANT: Kazuyasu NAKAMURA  
; APPLICANT: Kazuhisa UCHIDA  
; APPLICANT: Toyohide SHINKAWA  
; APPLICANT: Naoko YAMANE  
; APPLICANT: Motoo YAMASAKI  
; APPLICANT: No. US20030115614A1uo HANAI  
; TITLE OF INVENTION: ANTIBODY COMPOSITION-PRODUCING CELL  
; FILE REFERENCE: 249-202  
; CURRENT APPLICATION NUMBER: US/09/971,773  
; CURRENT FILING DATE: 2002-08-30  
; PRIOR APPLICATION NUMBER: JP 2000-308526  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/268,926  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Cricetus griseus  
US-09-971-773-23

Query Match 98.0%; Score 3024; DB 11; Length 575;  
Best Local Similarity 97.6%; Pred. No. 8.1e-266;  
Matches 561; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MRWGTGSRWIMLILFAWGTLILFYGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60  
Db 1 MRWGTGSRWIMLILFAWGTLILFYGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60

Qy 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHEILRRRIE 120  
Db 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHEILRRRIE 120  
Qy 121 NGAKELWFFLQSELKKLHLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Db 121 NGAKELWFFLQSELKKLHLEGNELQRADEILLDLGHHERSIMTDLVYLSQTDGAGDWR 180  
Qy 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCNKGCGYGCQLHHVYVCFMIAVGTQRT 240  
Db 181 EKEAKDLTELVRRTIYLNQPKDCSKARKLVNCNKGCGYGCQLHHVYVCFMIAVGTQRT 240  
Qy 241 LILESQNRWYATGCHWETVFRPVSCTDRSGLSGTHSGSEVNDKNIQVWELPIVDSLHPR 300  
Db 241 LILESQNRWYATGCHWETVFRPVSCTDRSGLSGTHSGSEVNDKNIQVWELPIVDSLHPR 300  
Qy 301 PYPPLAVPEDLADRLRLRHVHGDPAVMWVSQFVKYLIRPQPMLEKIEBEATKLGFKHPVI 360  
Db 301 PYPPLAVPEDLADRLRLRHVHGDPAVMWVSQFVKYLIRPQPMLEKIEBEATKLGFKHPVI 360  
Qy 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEAKTK 420  
Db 361 GVHVRTDKVGTAAAFHPIEEYVMVHVEHFQLLARRMQVKKRVYLATDDPTLLKEAKTK 420  
Qy 421 YSNYEFISDNTSWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Db 421 YSNYEFISDNTSWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
Qy 481 TLHPDASANFHLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Db 481 TLHPDASANFHLDDIYFVGQNAHQIAVYPHKPRTREEIPMEPGDIIGVAGNHWGYS 540  
Qy 541 KGINRKLGTGLPSYKVKREKIEVTKYPTYPEAEK 575  
Db 541 KGINRKLGTGLPSYKVKREKIEVTKYPTYPEAEK 575

## RESULT 3

US-09-839-136-10  
; Sequence 10, Application US/09839136  
; Patent No. US20020081694A1  
; GENERAL INFORMATION:  
; APPLICANT: Naoyuki TANIGUCHI et al.  
; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE  
; FILE REFERENCE: 2356-7  
; CURRENT APPLICATION NUMBER: US/09/839,136  
; CURRENT FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 09/442,629  
; PRIOR FILING DATE: 1999-11-18  
; PRIOR APPLICATION NUMBER: 08/913,805  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: PCT/JP97/00171  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 575  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-839-136-10

Query Match 97.0%; Score 2992; DB 9; Length 575;  
Best Local Similarity 96.2%; Pred. No. 6.6e-263;  
Matches 553; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MRWGTGSRWIMLILFAWGTLILFYGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60  
Db 1 MRWGTGSRWIMLILFAWGTLILFYGGHLVRDNDHPDHSRELSKILAKLERLKOQNEDL 60

Qy 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHEILRRRIE 120  
Db 61 RMAESLRIPGPIDQGTATGRVRLVEQLVKAKEQIENYKQARNGLGKDHEILRRRIE 120

```
QY 121 NGAKELWFFLOSELKKLKHLGEGNELQRAHDEILLDLGHHERSIMTDLYLSQTDGAGDWR 180
DB 121 NGAKELWFFLOSELKKLKHLGEGNELQRAHDEILLDLGHHERSIMTDLYLSQTDGAGDWR 180
QY 181 EKEAKDLTELQVRRITVYLNQPKDCSKARKLVNCNKGCGYGCQJHHVVYCFMIAYGQT 240
DB 181 EKEAKDLTELQVRRITVYLNQPKDCSKARKLVNCNKGCGYGCQJHHVVYCFMIAYGQT 240
QY 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
QY 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
QY 361 GVHVRRTDKVGTAAAFHPIEBEYMHVVEHFOLLARRMQVKKRYYLATDDPTLLKEATK 420
DB 361 GVHVRRTDKVGTAAAFHPIEBEYMHVVEHFOLLARRMQVKKRYYLATDDPTLLKEATK 420
QY 421 YSNYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
QY 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYPHKPRTEEEIPMEPGDIIIGVAGNHWGYS 540
DB 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYPHKPRTEEEIPMEPGDIIIGVAGNHWGYS 540
QY 541 KGINRKLGTGLYPSYKVRKEIETVKYPTYPEAK 575
DB 541 KGVNRKLGRGTGLYPSYKVRKEIETVKYPTYPEAK 575
```

## RESULT 4

US-09-839-136-2

; Sequence 2, Application US/09839136

; Patent No. US20020081694A1

; GENERAL INFORMATION:

; APPLICANT: Naoyuki TANIGUCHI et al.

; TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE

; FILE REFERENCE: 2356-7

; CURRENT APPLICATION NUMBER: US/09/839,136

; CURRENT FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: 09/442,629

; PRIOR FILING DATE: 1999-11-18

; PRIOR APPLICATION NUMBER: 08/913,805

; PRIOR FILING DATE: 1998-01-07

; PRIOR APPLICATION NUMBER: PCT/JP97/00171

; PRIOR FILING DATE: 1997-01-23

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 575

; TYPE: PRT

; ORGANISM: Pig

US-09-839-136-2

Query Match 94.1%; Score 2903; DB 9; Length 575;

Best Local Similarity 93.4%; Pred. No. 8.1e-255;

Matches 537; Conservative 14; Mismatches 24; Indels 0; Gaps 0;

```
QY 1 MRWTGSRWMTLILFANGTLLFYIGGHLVRNDHPDHSSELKILAKLERLKOQNEDL 60
DB 1 MRWTGSRWMTLILFANGTLLFYIGGHLVRNDHPDHSSELKILAKLERLKOQNEDL 60
QY 61 RMAESLRIPEGPIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHIEILRRRIE 120
DB 61 RMAESLRIPEGPIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHIEILRRRIE 120
QY 121 NGAKELWFFLOSELKKLKHLGEGNELQRAHDEILLDLGHHERSIMTDLYLSQTDGAGDWR 180
DB 121 NGAKELWFFLOSELKKLKHLGEGNELQRAHDEILLDLGHHERSIMTDLYLSQTDGAGDWR 180
```

```
QY 181 EKEAKDLTELQVRRITVYLNQPKDCSKARKLVNCNKGCGYGCQJHHVVYCFMIAYGQT 240
DB 181 EKEAKDLTELQVRRITVYLNQPKDCSKARKLVNCNKGCGYGCQJHHVVYCFMIAYGQT 240
QY 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
DB 241 LILESQNRWYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPR 300
QY 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
DB 301 PPYLPLAVPEDLADRLLRVHGDPAVWVVSQFVKYLIRPQWLEKEIEBEATKLGFKHPVI 360
QY 361 GVHVRRTDKVGTAAAFHPIEBEYMHVVEHFOLLARRMQVKKRYYLATDDPTLLKEATK 420
DB 361 GVHVRRTDKVGTAAAFHPIEBEYMHVVEHFOLLARRMQVKKRYYLATDDPTLLKEATK 420
QY 421 YSNYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
DB 421 YSNYEFISDINSISWSAGLHNRYTENSURGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480
QY 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYPHKPRTEEEIPMEPGDIIIGVAGNHWGYS 540
DB 481 TLHPDASANFHSLLDDIYFGQNAHNOIAVYPHKPRTEEEIPMEPGDIIIGVAGNHWGYS 540
QY 541 KGINRKLGTGLYPSYKVRKEIETVKYPTYPEAK 575
DB 541 KGVNRKLGRGTGLYPSYKVRKEIETVKYPTYPEAK 575
```

## RESULT 5

US-10-106-698-4658

; Sequence 4658, Application US/10106698

; Publication No. US20030109690A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides

; FILE REFERENCE: PA005P1

; CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 4658

; LENGTH: 515

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-106-698-4658

Query Match 86.2%; Score 2658; DB 15; Length 515;

Best Local Similarity 96.1%; Pred. No. 1.3e-232;

Matches 491; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

```
QY 65 ESURIPGPDIDQGTATGRVRLERQLVKAKEQIENYKQARNGLGKDHIEILRRRIENGAK 124
DB 5 KSLRIPEGPIDQGPAGRVRVLEEQLVKAKEQIENYKQARNGLGKDHIEILRRRIENGAK 64
QY 125 ELMPFLOSELKKLKHLGEGNELQRAHDEILLDLGHHERSIMTDLYLSQTDGAGDWRKEA 184
DB 65 ELMPFLOSELKKLKHLGEGNELQRAHDEILLDLGHHERSIMTDLYLSQTDGAGDWRKEA 124
QY 185 KDLTELQVRRITVYLNQPKDCSKARKLVNCNKGCGYGCQJHHVVYCFMIAYGQT 244
DB 125 KDLTELQVRRITVYLNQPKDCSKARKLVNCNKGCGYGCQJHHVVYCFMIAYGQT 184
QY 245 SONRYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPRPYYL 304
DB 185 SONRYATGWEVTFRVPVSETCTDRSGLSGTGHSGEVNDKNIQVVELPIVDSLHPRPYYL 244
```





;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 46107  
;; LENGTH: 81  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL109847.3  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
;; OTHER INFORMATION: SWISSPROT HIT: Q9VLZ7, EVALUE 6.40e+00  
;; OTHER INFORMATION: EST\_HUMAN HIT: AW002060.1, EVALUE 5.00e-43  
US-09-864-761-46107

Query Match 13.8%; Score 426; DB 9; Length 81;  
Best Local Similarity 96.3%; Pred. No. 5.8e-31;  
Matches 78; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 280 EVNKNQVVELPIVDSLHPPYPLAVPDLADRLRLRVHGDPAVWVSQFVKYLIRPQ 339  
Db 1 EVKKNQVVELPIVDSLHPPYPLAVPDLADRLRVHGDPAVWVSQFVKYLIRPQ 60  
  
Qy 340 PWLEKEIEATKLGFKHPVI 360  
Db 61 PWLEKEIEATKLGFKHPVI 81

RESULT 8  
US-10-369-493-12504  
;; Sequence 12504, Application US/10369493  
;; Publication No. US20030233675A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Cao, Yongwei  
;; APPLICANT: Hinkle, Gregory J.  
;; APPLICANT: Slater, Steven C.  
;; APPLICANT: Goldman, Barry S.  
;; APPLICANT: Chen, Xianfeng  
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
;; FILE REFERENCE: 38-10(52052)B  
;; CURRENT APPLICATION NUMBER: US/10/369,493  
;; PRIOR FILING DATE: 2003-02-28  
;; PRIOR APPLICATION NUMBER: US 60/360,039  
;; PRIOR FILING DATE: 2002-02-21  
;; NUMBER OF SEQ ID NOS: 47374  
;; SEQ ID NO 12504  
;; LENGTH: 1055  
;; TYPE: PRT  
;; ORGANISM: Aspergillus nidulans  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (1)-(1055)  
;; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-12504

Query Match 4.1%; Score 127; DB 12; Length 1055;  
Best Local Similarity 25.6%; Pred. No. 0.041;  
Matches 44; Conservative 38; Mismatches 60; Indels 30; Gaps 7;  
  
Qy 42 ELSKILAKLERLKQNEDLRMAESLRIPGPIQGTATGRVRLVEQLVKAKEQIENYK 101  
Db 771 KIQKIMIBIDSAKRSLADARRVQELSAHKP--SKTDSARVKVLEEQIVGLEQIEDLR 828

Qy 102 KOARNGLGKDHKILRRRI-----ENGAKELWFFLOSEL-----KKLKH 139  
Db 829 SQ-KGGIEEIQALQNKIMEVGGVRLRSQAKVDGLKEQISLLSEISNAEVARSKNEKL 887  
  
Qy 140 LEGNELOR-HADHILLDGHHSIMTDLVYLSQTDGAGDWREK--EAKDLT 188  
Db 888 IKKHENARAFAEKELANVAEELQKLNEDV--ANOANDASGWKEKVDQAQDVS 937  
  
RESULT 9  
US-10-104-047-3419  
;; Sequence 3419, Application US/10104047  
;; Publication No. US20030236392A1  
;; GENERAL INFORMATION:  
;; APPLICANT: HELIX RESEARCH INSTITUTE  
;; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
;; FILE REFERENCE: HI-A0105  
;; CURRENT APPLICATION NUMBER: US/10/104,047  
;; CURRENT FILING DATE: 2002-03-25  
;; PRIOR APPLICATION NUMBER:  
;; PRIOR FILING DATE:  
;; NUMBER OF SEQ ID NOS: 4096  
;; SOFTWARE: PatentIn ver. 2.1  
;; SEQ ID NO 3419  
;; LENGTH: 485  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-104-047-3419

Query Match 4.0%; Score 123; DB 12; Length 485;  
Best Local Similarity 20.3%; Pred. No. 0.03;  
Matches 92; Conservative 70; Mismatches 140; Indels 152; Gaps 23;  
  
Qy 45 KILAKLERLK-QNEDLRMAESLRIPGPIQGTATGRVRLVEQLVKAKEQIENYKQ 103  
Db 13 KKVTKAERLKLQEEERLKEE-----EASLK-----YEKE 45  
  
Qy 104 ARNGLGKDHKILRRRIENGAKELWFFLOSELKKLHLEGNELORHADI-----LLDLG-- 157  
Db 46 -----EMERLEIQRIEKEW-----HRLKADLERNEELEYLLERCFF 86  
  
Qy 158 -----HHSIMTDLVYLSQTDGAGD-----WREKBAKOLTELVR----- 193  
Db 87 EAEKLGKETKLLSQWKHYIQDGSPPDPSVAQEMTFTSLWKEKTNETFEEVIEKSKVLN 146  
  
Qy 194 -----RITYLQNPXDCSKARKLVNINKGCGYQGLHHVYCFMIAVGTORTLLESQNW 248  
Db 147 LIEKLFILLETLP-PCDLQDNIIQYQESI---LQLOQLLH---LKFQVATEILLKQ--- 196  
  
Qy 249 RYATGGWETVFRPVSETCDRSGLSTGHSVNDKNIQVVLELPVDSLHPRPPYPLAV 308  
Db 197 -----ASTLAD---LDSGMEKVIKDN-----VTLVYWANLKNPRHRSVRF 236  
  
Qy 309 PE-----DLADRLRLVHGD-----PAVWV---VSQFVKYLIR-PQP 340  
Db 237 SETQIGFEIPRILATSDIAVRLHLLTHYDVSALHPVSTPSPKEYTSVAVTELKVDVKVNEK 296  
  
Qy 341 WLEKIEEATKLGFKHPVIGVHVVRTDKVGTAAFPHEIPEYVHVVEHFOLL--ARMQ 398  
Db 297 AISKEVEESKQ-----QERGSHLIOEEETKVEEQGDIEVKMSSABEESAICERBMK 351  
  
Qy 399 VDKRVVYLATDDPTLLKAEATKYSNYEFISDNI 432  
Db 352 VLSETVSAA---QLLLVENSE--KPDFFDNVV 380  
  
RESULT 10  
US-09-839-136-12  
;; Sequence 12, Application US/09839136  
;; Patent No. US20020081694A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Naoyuki TANIGUCHI et al.  
;; TITLE OF INVENTION: ALFA 1-6 FUCOSYLTRANSFERASE

```
; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
; US-09-839-136-12

Query Match      4.0%; Score 122; DB 9; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.00043;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 352 KLGFKHPVIGVHVRTRDKVGTAAAF 376
DB 1 KLGFKHPVIGVHVRTRDKVGTCAAF 25

RESULT 11
US-10-117-937-596
; Sequence 596, Application US/10117937
; Publication No. US20030220239A1
; GENERAL INFORMATION:
; APPLICANT: CTL IMMUNO THERAPIES CORP.
; APPLICANT: SIMARD, John, J.L.
; APPLICANT: DIAMOND, David, C.
; APPLICANT: LIU, Liping
; APPLICANT: XIE, Zhidong
; TITLE OF INVENTION: EPITOPE SEQUENCES
; FILE REFERENCE: CTLIMM.027A
; CURRENT APPLICATION NUMBER: US/10/117,937
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/282,211
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/337,017
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/363,210
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 602
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 596
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-117-937-596

Query Match      3.8%; Score 118.5; DB 12; Length 976;
Best Local Similarity 19.5%; Pred. No. 0.22;
Matches 91; Conservative 66; Mismatches 171; Indels 139; Gaps 18;

QY 45 KILAKLERLUQNEDLRMAESLRIPGIDQGTATGRV-RVLEBQLVKAKEQIENYKKQ 103
DB 157 KVSLEEGIQENKDL-----IKENNATRLCNLLKETCARSAEKKYKYE 203

QY 104 -----ARNGLCK---DHEILRRRIENKAKELWFFLOSELKKLKHLEGNLQRHADBIIDL 144
DB 204 RRETRQVMDLANNIEKMTATGEGLRVQAEENRLEMHFKLKDYEKIOHLEQYKKEIND 263

QY 145 LQRHADEIILLGHHSRIMTDLYL--SOTDAGDWRE-----K 182
DB 264 KEQVSLLIQITKENK-MKDLTFLEESRDKNVQLEKTKLOSENKQSTIEKQHHLTK 322

QY 183 EAKDLTELVRRI-TYLNQPKDCSKARKLVNCNINCGGCGQLHHVYCFMAYGTQRTL 241

; FILE REFERENCE: 2356-7
; CURRENT APPLICATION NUMBER: US/09/839,136
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/442,629
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 08/913,805
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: PCT/JP97/00171
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
; US-09-839-136-12

Query Match      4.0%; Score 122; DB 9; Length 25;
Best Local Similarity 96.0%; Pred. No. 0.00043;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 352 KLGFKHPVIGVHVRTRDKVGTAAAF 376
DB 1 KLGFKHPVIGVHVRTRDKVGTCAAF 25

RESULT 11
US-10-117-937-596
; Sequence 596, Application US/10117937
; Publication No. US20030220239A1
; GENERAL INFORMATION:
; APPLICANT: CTL IMMUNO THERAPIES CORP.
; APPLICANT: SIMARD, John, J.L.
; APPLICANT: DIAMOND, David, C.
; APPLICANT: LIU, Liping
; APPLICANT: XIE, Zhidong
; TITLE OF INVENTION: EPITOPE SEQUENCES
; FILE REFERENCE: CTLIMM.027A
; CURRENT APPLICATION NUMBER: US/10/117,937
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/282,211
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/337,017
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/363,210
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 602
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 596
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-117-937-596

Query Match      3.8%; Score 116.5; DB 12; Length 919;
Best Local Similarity 22.5%; Pred. No. 0.3;
Matches 108; Conservative 70; Mismatches 160; Indels 143; Gaps 25;

QY 50 LERLQONEDLRMAESLRIPGIDQGTATGRVYRVEQLVKAKEQIENYKKQARNGLG 109
DB 522 LSRLNQLLEGLREL--GFQTP-----DLAKAFQKLRLRLERLEURLKLENSLEEKVRN-LS 575

QY 110 KDHEILR-----RRI---ENGAKELWFFLOSELKKLKHLEGNLQRHADBIIDL 156
DB 576 REEVALREAKTRALEVLQRIKKEEBAEKLTLSSESKLERM-----LVSKAEDLATRL 631

QY 157 GHHSRIMTDLY-----YLSQTD---GAGDWREKAKDL-----TELVRRIYTL 198
DB 632 GITAVRSDDLLEKAREALEGVDKELSAIERREARRLEKAAKLKWEAEQVMKRLLEL 691

QY 199 QNPDKCSKARKLVNCNINCGGCGQLHHVYCFMAYGTQRTLI-LESQWRYATGGWET 257
DB 692 E-----ABEKKLRKEVSRKSEIARLKEV-----QNTLAELDDRTISR----- 728

QY 258 VFRPVSETCTDRSGLSGTHSGSEVNDKNIQVVELPIVDSLHPRPPYPLA----VPEDLA 313
DB 729 IDRMGELQTRIREMKSRKASGE-----EALKL-----YLPAAASRRIMEIG 771

QY 314 D-----RLLRVHGDPAVWVVSQFVKYLRPQWPLEKEIEATKLGFKHPVIGVHVR 365
DB 772 EIAVRRLLAV-----LEDMNDILSR--FNLDVAGVIREKAA 807

QY 366 ---RTDKVGTAAAFPIEENYVHVEHFQ-----LARRMOVDKKRVYLAATDDPTLLKEA 417
```

Db 808 REIEVKAIGNGAYRPLE--AVSGGERTVLAISFVLALNKAVGGKLGFLALDEPTANLDE 865  
Qy 418 KTKYNGYFISDINSWSAGLHNRRTENSRLGVLDIHFLSQADF--LVC--TFSSOVCR 473  
Db 866 DRRRLVEVRGISV-----EGLVRLVVVTHHEDVRDYADTICLVTRTQOQSR 914  
Qy 474 V 474  
Db 915 V 915  
RESULT 13  
US-10-369-493-6527  
; Sequence 6527, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIORITY FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6527  
; LENGTH: 1992  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6527

Query Match 3.8%; Score 116.5; DB 12; Length 1992;  
Best Local Similarity 19.1%; Pred. No. 0.97;  
Matches 89; Conservative 69; Mismatches 148; Indels 161; Gaps 19;  
Qy 42 ELKILAKLERLKOQNEDLRRMAE-----SURPEGIDOGTATGRVVRV 85  
Db 954 KLSDTGQLEDQMOERNEDLARQKKTQDLSDTKKHVQDLESLFKABQEQKQSRDHNR 1013  
Qy 86 LEEQLVAKAEQIENYKQARNGKGLKDEHILRRRIENGAKELWFFLOSELKHLKLE--GN 143  
Db 1014 LQDENANQDEAVAKLNKE-----KKHQ-----EESNRKLNEDLQSEEDKVNHLKTRN 1061  
Qy 144 ELQRHAEILLDLGHHERSIMTDLYLSQTDGAGDWREKEAKDLTELVRITTYLQNPKD 203  
Db 1062 KLEQQMDEL-----EENID-----REKSRGDIKAKRV-----EGD 1094  
Qy 204 CSKARKLVNINCKGCGQLHHVVCYPMIAYGTORTLILESQNRWRYATGWETVFRPVS 263  
Db 1095 LKVAQENIDEITK-----QKHQV-----R 1112  
Qy 264 ETCTDRSGLSCHWSGEVNDKNIQVWELP-IVDSLHPRPPYPLAVPDEL-ADRLLRVHG 321  
Db 1113 ETTLKRKEEDLHHTTAKLAENNSIIAKLQRLKELTARNAEI-----EELAEARNRQKS 1168  
Qy 322 DPAVWVWSQFVKYLRPPQWLEKEETEATKGLGPKHPVIGVHV-----R 365  
Db 1169 D-----RSRSEARELEELTERLEQOOGGATAQLEANKKREAEIAKLREK 1214  
Qy 366 RTDKVGTAAFPPIE-----EYVHVVEHFQLLAR-----RMQVDKRVVYLATDDP 411  
Db 1215 EEDSLNHETAISSLKRKHGDSVAELTEQLETLQKLKAKSEAKSKLQRLDEESQHAATDSE 1274  
Qy 412 T-----LLKEATKYSNTEFISD-----NSISWSAGLHNRV--TENS 446  
Db 1275 VRSRDLEBKALTKTIEVQYSELQTKADEQSRQLQDFPAALKNRNNNS 1321

RESULT 14  
US-10-369-493-5427  
; Sequence 5427, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIORITY FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5427  
; LENGTH: 678  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5427

Query Match 3.8%; Score 116; DB 12; Length 678;  
Best Local Similarity 18.1%; Pred. No. 0.21;  
Matches 86; Conservative 81; Mismatches 185; Indels 122; Gaps 16;  
Qy 38 HSRRELSKILAKLERLKOQNEDLRRMAESLRIPGPDQGTATGRVVRVLEBOLVKAKEQI 97  
Db 255 HLGDMDSIELEFASKMEQADEMRSLABEV-----FNERRAKEBELQGLEIEV 302  
Qy 98 ENYKQARNGKGLKDEHILRRRIENGAKELWFFLOSELKHLK----- 138  
Db 303 EEQKKL-----NEAVTHAMDPMQKEKYEDLSEAKLLRERVVMEAKNEDLDDRISK 354  
Qy 139 ---HLEGNELQFHA---DEILLDLGHHERSIMTDLYLSQTDGAGDWREKEAKDLTELVRV 192  
Db 355 YEIEIRSNPLKKAIAQLOETLTLKKQBEKLMEDMQSALTPEA---WRDKSENKMKQLNA 411  
Qy 193 RRTYVQNPQ-----DCSKARKLVNINCKGCGQLHHVVCYPMIAYGTQ-----R 239  
Db 412 DLVVIKQHKTKVQDIQISLASELHEYDSQGEAQIMAHHTKYLDLILSKSTMLDDTTTENYPQ 471  
Qy 240 TLILESQNRWRYATGWETVFRPVS-----TCTDRSGLSCHWSGEVNDK----- 284  
Db 472 QIVYQDDIEFSDAVVLILRKISANLKKVNLDEQITDLDERGLTL--OTGNVDELKEMH 529  
Qy 285 -----NIQVWELPI---VDSLHPRPPYPLAVPDELADRLLRVHGDPVAVWVWSQFVK 333  
Db 530 VRLQEEELISIEFMELALNEEDINLETEK-----KIDQELAGVGNVDS----- 574  
Qy 334 YLIRPQPWLEKIEBEATKGLGPKHPVIGVHVRRTDK--VGTEAAPHPIEYVHVVEHFQ 391  
Db 575 -----GLRQLEERQKLEDEAPARSHENQOLEANVASIRNELHSPGYSQH-----K 622  
Qy 392 LLARRMQVDKRVVYLATDDPDLLEAKTKYSNYEFISDINSWSAGLHNRRTEN 445  
Db 623 MLRERLEAVEKRTAAKSLDMSLRKTE-----IDYEDIKTESIRLOQBEYNNMLLTN 672

RESULT 15  
US-10-205-823-273  
; Sequence 273, Application US/10205823  
; Publication No. US20030308963A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarepu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi

```
; APPLICANT: Womsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 273
; LENGTH: 1285
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-273

Query Match          3.8%; Score 116; DB 15; Length 1285;
Best Local Similarity 19.7%; Pred. No. 0.56;
Matches 71; Conservative 51; Mismatches 130; Indels 108; Gaps 15;

QY 39 SSRELSKILAKLERLKQONEDLRRMAESLRIPGDPIDQGTATGRVRLVEEQLVKAKEQIE 98
Db 905 SSELLSALQKKQOEAEERLRIOEME-----KERREREDEKRRKKEBEE 953
QY 99 NYKQARNGLGKDHEITLRRRIENGAKELWFFLOSELKCLKHLEGN-----ELQRHADEIL 153
Db 954 RRMKLEMEARKQOEERKKEDEDEKRIQAEVEAQLARQKEEESQQAQVLEQERRDRELA 1013
QY 154 LDIGHHERSITMDL-----YLSQTDGAGD--WREKAKDLTELVRRTITLQNPXDCSKA 207
Db 1014 LRIAQSEAEILSDEAQADLALRRNDGTRPMTPEQNAKEMSEFLSRGPAVL-----ATKA 1068
QY 208 RKLVCNINKGCGYGCQLHHVYCFMIAYGTQRTLILESQNWRYATGGWETVFRPVSETCT 267
Db 1069 -----AAGTKK--YDLSKWKYAE-----LRDTINTSC 1093
QY 268 D-----RSLGSLTGH-WSGEVNDKNIQ-----VVELPIVDSLHPRPPYPLPLA 307
Db 1094 DIELLAACREEFHRRLLKVYHAWKSKNKNKNTETEQRAPKSVTDYDFAPFLNNSPQONPAA 1153
QY 308 -----VPEDLADRLLV-----HGDPVAV---WVVSQFVKYLIRPOPWLEKETE 347
Db 1154 QIPAROREIEMNRQRFPIRIPADQYKDPQSKKKGWYAHF-----DGPWTARQME 1207
```

Search completed: February 2, 2004, 08:52:04  
Job time : 218 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 17 Seconds  
(without alignments)  
3252.764 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRAWTGSRWIMLILFANGT.....YKREKIETVKYPTPEAEK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3003	97.3	575	1 JC5432	glycoprotein 6- $\alpha$ p
2	982	31.8	818	2 T08621	hypothetical prote
3	126	4.1	946	2 S28061	SCPI protein - rat
4	125.5	4.1	845	2 I48176	synaptonemal compl
5	125	4.1	1025	2 H86250	hypothetical prote
6	118	3.8	1173	2 T43527	sp8 protein - firs
7	118	3.8	1727	2 T50073	myosin-like coiled
8	117	3.8	1744	2 F86161	Fl003.10 protein -
9	117	3.8	1871	2 D96796	probable heat shoc
10	117	3.8	1955	2 T30934	myosin-like protei
11	116.5	3.8	919	2 B72765	hypothetical prote
12	116.5	3.8	1992	1 S02771	myosin heavy chain
13	116	3.8	678	2 H88187	protein C18H9.8 li
14	115.5	3.7	1427	2 S22695	restin - human
15	115	3.7	1137	2 T19414	hypothetical prote
16	115	3.7	1156	2 B70356	chromosome assembl
17	115	3.7	1325	2 S16129	dynein-associated
18	114	3.7	755	2 T34567	hypothetical prote
19	113.5	3.7	1011	2 T13055	dynamain associated
20	113.5	3.7	1023	2 T33338	hypothetical prote
21	113.5	3.7	2020	2 T21174	hypothetical prote
22	113	3.7	1094	2 T13053	dynamain associated
23	112.5	3.6	832	2 H72278	alpha-mannosidase-
24	112.5	3.6	1270	2 T09194	adaptor protein in
25	112.5	3.6	2007	1 B43402	myosin heavy chain
26	111.5	3.6	1281	2 JCS368	dynactin 1 - mouse
27	111.5	3.6	1392	2 A43336	microtubule-vesicl
28	111.5	3.6	1935	1 A37102	myosin beta heavy
29	111	3.6	993	2 S49461	synaptonemal compl

30	111	3.6	1397	2 S68420	citron - mouse
31	111	3.6	2142	2 T08621	centrosome associa
32	110.5	3.6	542	2 E86496	hypothetical prote
33	110.5	3.6	542	2 F81543	hypothetical prote
34	110.5	3.6	978	2 A70387	conserved hypothet
35	110	3.6	581	2 B87768	protein F54C1.1 li
36	110	3.6	1057	2 F89892	carbamoyl-phosphat
37	109.5	3.5	542	2 F72127	hypothetical prote
38	109.5	3.5	1179	2 F71190	probable chromosom
39	109.5	3.5	1282	2 JF0120	glycoprotein A - m
40	109.5	3.5	1935	1 S06006	myosin beta heavy
41	109	3.5	589	2 T48905	gamma response 1 p
42	109	3.5	1110	2 T49091	myosin heavy chain
43	109	3.5	1934	2 I48153	protein F55A8.2 li
44	108.5	3.5	521	2 D88640	conserved hypothet
45	108	3.5	380	2 F75103	

ALIGNMENTS

RESULT 1

JC5432  
glycoprotein 6- $\alpha$ -L-fucosyltransferase (EC 2.4.1.68) precursor - human  
N:Alternate names: alpha 1-6 FucT; GDP-L-fucose:N-acetyl-beta-D-glucosaminyl:6- $\alpha$ -L-f  
C:Species: Homo sapiens (man)  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Jun-2000  
C:Accession: JC5432; PC4322  
R:Yanagidani, S.; Uozumi, N.; Ihara, Y.; Miyoshi, E.; Yamaguchi, N.; Taniguchi, N.  
J. Biochem. 121, 626-632, 1997  
A:Title: Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha  
A:Reference number: JC5433; MUID:97279058; PMID:9133635  
A:Accession: JC5432  
A:Molecule type: mRNA  
A:Residues: 1-575 <YAN1>  
A:Cross-references: DDBJ:D89289; NID:G2055306; PIDN:BAA19764.1; PID:G2055307  
A:Accession: PC4322  
A:Molecule type: protein  
A:Residues: 68-87/352-376;419-432 <YAN2>  
C:Comment: This enzyme catalyzes the transfer of fucose from GDP-fucopyranoside to aspar

C:Genetics:  
A:Gene: GDB:FUT8  
A:Cross-references: GDB:9786294; OMIM:602589  
A:Map position: 14q23-14q23  
C:Superfamily: human glycoprotein 6- $\alpha$ -L-fucosyltransferase  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1-22/Domain: signal sequence #status predicted <Sig>  
F:23-575/Product: glycoprotein 6- $\alpha$ -L-fucosyltransferase #status predicted <MAT>

Query Match		97.3%	Score 3003;	DB 1;	Length 575;
Best Local Similarity		96.5%	Pred. No. 1.6e-194;	Mismatches 9;	Indels 0; Gaps 0;
Matches 555;		Conservative 11;			
QY	1	MRAWTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKQONEDL	60		
DB	1	MRPWTGSRWIMLILFANGTLLFYIGGHLVRDNDHPDHSRELSKILAKLERLKQONEDL	60		
QY	61	RMMAESLRIPSPIDQGTATGVRVLEQLVKAEIQENYKQARNGLGKDHILRRRIE	120		
DB	61	RMMAESLRIPSPIDQGTATGVRVLEQLVKAEIQENYKQARNGLGKDHILRRRIE	120		
QY	121	NGAKELWFFLOSELKLLKLGNELOHRADELLDLGHHERSIMTDLYLSQTDGAGDWR	180		
DB	121	NGAKELWFFLOSELKLLKLGNELOHRADELLDLGHHERSIMTDLYLSQTDGAGDWR	180		
QY	181	EKEAKDITELVQRITTYLQNPDKCSKARKLVNCKGCGYQLHHVVCYFMIAYGTORT	240		
DB	181	EKEAKDITELVQRITTYLQNPDKCSKARKLVNCKGCGYQLHHVVCYFMIAYGTORT	240		
QY	241	LILSQNWRYAFGGWETVFRPVSETCTDRSGLSGTHWSGEVNDKNIQVLEPIVDSLHPR	300		
DB	241	LILSQNWRYAFGGWETVFRPVSETCTDRSGLSGTHWSGEVNDKNIQVLEPIVDSLHPR	300		

QY 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQPWLEKIEEATKLGKHPVI 360  
DB 301 PPYLPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQPWLEKIEEATKLGKHPVI 360  
QY 361 GVHVRRTDKVGTAAFPHPTEEVYVHVEHFQLLARRMQVVKRVYLATDDPTLLKEATK 420  
DB 361 GVHVRRTDKVGTAAFPHPTEEVYVHVEHFQLLARRMQVVKRVYLATDDPTLLKEATK 420  
QY 421 YSNYEFISDNSISWSAGLHNRNTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
DB 421 YSNYEFISDNSISWSAGLHNRNTENSRLGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
QY 481 TLHPDASANFHSDDIYIFGGQNAHNQIAVYHPKRTBEIEIPMEPGDIIIGVAGNHWGYS 540  
DB 481 TLHPDASANFHSDDIYIFGGQNAHNQIAVYHPKRTBEIEIPMEPGDIIIGVAGNHWGYS 540  
QY 541 KGINKLGTGLYPSYKVEKIEETVKYTPYPAEK 575  
DB 541 KGINKLGTGLYPSYKVEKIEETVKYTPYPAEK 575

RESULT 2  
T32154  
hypothetical protein C10F3.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
R:Accession: T32154  
R:Davidson, S.; Wohlmann, P.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid C10F3.  
A:Reference number: Z21127  
A:Accession: T32154  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-818 <DAV>  
A:Cross-references: EMBL:AF022968; PIDN:AAB69888.1; GSPDB:GN00023; CBSP:C10F3.6  
A:Experimental source: strain Bristol N2; clone C10F3  
C:Genetics:  
A:Gene: CBSP:C10F3.6  
A:Map position: 5  
A:Introns: 30/1; 72/2; 235/2; 258/3; 285/2; 315/3; 457/3; 491/1; 544/3; 556/1; 626/1; 64

Query Match 31.8%; Score 982; DB 2; Length 818;  
Best Local Similarity 36.8%; Pred. No. 3.4e-58;  
Matches 211; Conservative 122; Mismatches 191; Indels 50; Gaps 11;  
QY 3 AWTGSRWIMLILFAWGTLLFYIGHLVRDNDHPDHSRELKILAKLERLQQNEDLRR 62  
DB 6 AAVGTVMVMTMFLYSQL-----SNNOSGDSIRAWRQTKEADIKLOEQNEDLKS 56  
QY 63 MAESLRIPGEPIDQGTATGRVRLER-----QLVKAKEQIENYKQARNGLGRD 111  
DB 57 ILEKER-----QERNDQHKIMEQSHQLPNPNPNSLPKPEPVKEIISKSILG-PVQ 108  
QY 112 HEILRRRIENGAKELWFFLOSELKKLKHLEGNLQRADEIILDLGHHSRIMTDLYLS 171  
DB 109 QEVQKMLDDIREMFYLLHSQIE-----NSTKILLET--QMISLMGLSAQLE 155  
QY 172 QTDGADWREKAKDITELVQRITVYLOPKDCSKARKLVNCKGCGCOLHHVVYCF 231  
DB 156 KLEGSEERFKORTAITQIRFIKSEKLOPKACSEAKTLVNCIDKECGCOLHHVYCA 215  
QY 232 MIAYGTQRTLIL--ESQNRWYATGWTFRPVSCTDRSGLSGTGHWSE--VNDKNIQ 287  
DB 216 ITAFATQRMVWLKRDSSWKYSHGWTVPFKLSKCSFDE--AVGNTAKFPFAPSPAR 272  
QY 288 VVELPIVDSLPRPPYPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPQPWLEKETE 347  
DB 273 VVSLGIVDSLITKPTFLPOAVPEQLLESLSLHSHPPAFFVGTFTSYLMRFSNATQEKLD 332  
QY 348 EATKLGKFP-KHPVIGVHVRTDKVGTAAFPHPTEEVYVHVEHFQLLARRM-OVDKRVY 405  
DB 333 KALKSIPLDKGPVIGLQIRTKDVKGTAAFPALHAKYEMTEIWFVKVEKROGKPLERRIF 392

QY 406 LATDPTLLKEATKYSNYEFISDNSISWSAGLHNRNTENSRLGVILDIHFLSQADFLVC 465  
DB 393 IASDDPTVPPAKNDYPNYEVGYSTEAIAQLANNRYTDASLMGVITDIYILSKVNYLVC 452  
QY 466 TFSSQVCRVAYEIMQTLHPDASANFHSDDIYIFGGQNAHNQIAVYHPKRTBEIEIPMEP 525  
DB 453 TFSSQVCRMGVELRQPSGADGSKFHSDDIYIFGGQNAHNQIAVYHPKRTBEIEIPMEP 512  
QY 526 GDIIGVAGNHWGYSKGINKLGTGLYPSYKVR 559  
DB 513 GDKVGIAGNHWGYSKGTNRQTYKEGVFPYKIR 546

RESULT 3  
S28061  
SCP1 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
R:Accession: S28061  
R:Meuwissen, R.L.J.; Offenberger, H.H.; Dietrich, A.J.J.; Rieswijk, A.; van Iersel, M.;  
EMBO J. 11, 5091-5100, 1992  
A:Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase  
A:Reference number: S28061; MUID:93099884; PMID:1464329  
A:Accession: S28061  
A:Molecule type: mRNA  
A:Residues: 1-946 <MEU>  
A:Cross-references: EMBL:X67805; NID:G57212; PIDN:CAA48006.1; PID:G57213  
C:Genetics:  
A:Gene: SCP1  
C:Keywords: DNA binding

Query Match 4.1%; Score 126; DB 2; Length 946;  
Best Local Similarity 20.6%; Pred. No. 1.9;  
Matches 109; Conservative 78; Mismatches 180; Indels 162; Gaps 24;  
QY 42 ELSKILAKLERLQQNEDLRRMAESLRIPGEPIDQGTATGR-VRVLEQLVKAKEQIENY 100  
DB 104 ENKVSLKLEEEIOENKDL-----IKENNATRWCNLLKTCARSABKTSKY 150  
QY 101 KKQ-----ARNGLGK--DHEILRRRIENGAKELWFFLOSELKKLKHLEGNELQ 147  
DB 151 EYEREETQVYVDLNNNIEKMLAFEEELRVQAEARLEWHFKLKEDHEKIQHLE-EYQK 209  
QY 148 HAD-----EILLDLGHHSRIMTDLYL--SQTGADWREKAKDITELVQRITVY 199  
DB 210 EVNNKENQVSLLIQSTETENKMKDLTLLBESRDKANQLEBK-----TKLDENLKELN 264  
QY 200 NPKDCSKARKLVNCKGCGCOLHHVVYCFMAYGTQRTLLBSSQNRWYATGCGWTVF 259  
DB 265 EKKDHLTS-----ELEDIKMSQVRSMSQTKLT---EEDLQIAT---KTIY 303  
QY 260 RPVSS-----TCTDRSGLSGTGHWSEVND-----283  
DB 304 QLTEEKBAEMLNKAETHSLVTELKATCTLEELRTEQORLENEDQLKLTMLQ 363  
QY 284 -----KNIQVVELPIVDSLPRPPYPLAVPEDLADRLRLRVHGDPAVWVVSQ 330  
DB 364 KKSSELEEMTKFNKKEVELEBELKTI-----LAEDQKLLDEKKQVEKLABELQKBEQ 415  
QY 331 FVKYLIRPQPWLEKIEEATKLGKFPVIGVHVRTDKVGTAAFPHPTEEVYVHVEE-- 388  
DB 416 ELTPILQTR---EKEIHDL-----VQVTVTKTSE-----EHLVKQVEEMK 453  
QY 389 -----HFQLLARR--MOVDKRVYLATDDPTLLKEAKTKYSNYEFISDNSISWSA 436  
DB 454 TELEKEKLNIELTANSMDLLENKLVQESADMDVLELKGHQEDIIINCKQBERMLKQLE 513  
QY 437 GLHNRNTENSRLGVILDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480  
DB 514 TLEEK--EMNLRDELESVRKESFIQOGDEVKCKLDKSEENARSIEYEVLK 560

## RESULT 4

I48176  
synaptonemal complex protein - golden hamster (fragment)  
C:Species: Mesocricetus auratus (golden hamster)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: I48176  
R:Doelson, M.J.; Pearlman, R.E.; Karaiskakis, A.; Spyropoulos, B.; Moens, P.B.  
J. Cell Sci. 107, 2749-2760, 1994  
A:Title: Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction  
A:Reference number: I48176; MUID:95181577; PMID:7876343  
A:Accession: I48176  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-845 <RES>  
A:Cross-references: GB:L32978; NID:g598124; PIDN:AAC42039.1; PID:g598125  
C:Genetics:  
A:Gene: synl

Query Match 4.1%; Score 125.5; DB 2; Length 845;  
Best Local Similarity 19.7%; Pred. No. 1.7;  
Matches 85; Conservative 81; Mismatches 161; Indels 105; Gaps 17;  
QY 42 ELSKILAKLERLKQNEQLRMAESLRIPGPID--QGTATGRVRLVEQLVKAKEQIEN 99  
DB ||||| : : : : : ||| : : : : : ||||| : : : : : ||||| : : : : :  
288 ELKTLAEDQKLLDEKQVKEKLAELQKEQELTLLQTRKEVHDLLEQLLVTKISDN 347  
QY 100 YKQARNGLGKDHETLRRRIENGAKELWFFLQSELKLLKH-----LEGNELQRH 148  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
348 YSKQV-----EELTKLEE-----EKLKNAELTASCGLSLENNKLTQE 386  
QY 149 ADEIILLDGHHERSMT----DLYLSQTDGAGWREKEAKDLTELVRITLYLONPKDC 204  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
387 TNDMALEKKYQEDITNSKKQERMLKQIENLEE-KETHLDELSEVRKEFIQOGNEVKC 445  
QY 205 -----SKAKLVNINCGGCGVGCQHLHVYCFMAYGTQRTLLILESO--NWRVATGWE 256  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
446 KLDKSEENARSTCEVLK-----REKQMKILENKNLTKQA---E 483  
QY 257 TVFRPVSECTDRSL--STGHSGEVNDKNIQVVELPI-VDSLPR-----PPVPLAV 308  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
484 NKSYIEBELHQENKALKKSSAESQNLNAYEIKVKNKQLELSAKQKQFQMTDNQKEIE 543  
QY 309 PEDLADRLRHVHGDPAVMVSVFVYLRPQWLEKEIE-EATKLGFKHPVIGVHVRR 367  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
544 VKKISEKLLGLGEVAKAWDAVK-----LQKEIDLRCQHKIAENVALMEKHQY 595  
QY 368 DKVGTG-----AAPHIEEYVHVEHFQLLARMQV---DKKRVYLAT 408  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
596 DRIVEERSELGLCKNRQEQLSVKTALTELSNIRNELSVLKKQLEIERBEKELKLEK 655  
QY 409 DPDTLLKEAKTK 420  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
656 ENTALIKKKDK 667

## RESULT 5

H86250  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86250  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H86250

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1025 <STO>  
A:Cross-references: GB:ASC05172; NID:g4835785; PIDN:AAD30251.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 4.1%; Score 125; DB 2; Length 1025;  
Best Local Similarity 17.5%; Pred. No. 2.4;  
Matches 105; Conservative 84; Mismatches 207; Indels 204; Gaps 27;  
QY 5 TGSWRWMLILFANGTLLF-YI-----GHLVRDNDHPDHSRELSKILAKLERLQKQ 56  
DB ||||| : : : : : ||| : : : : : ||||| : : : : : ||||| : : : : :  
362 SGNN-WFAEVVPGCALVDWVADGPPKGAFLDNNNGYQDFHALVPQKLPPELYWLEEE 420  
QY 57 NEDLRRMAESLFIPEGPI-----DQGTATGRVRLVEQLVKAKEQIENYKQARNGLG 109  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
421 NMI FRKLOEDRELKEEVNRAKMEKTARLKAETKRTLKKFLLSQKDVV--YTEPLEIQAG 478  
QY 110 KDHEIL---REFIENGAKELWFFLQSELKLLKHLEG-----NELQRHA----- 149  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
479 NPVTLYNPANTVNLNGKEVWF--RGSFNRWTHRLGPLPPQKMEATDDESSHVKTTAKVP 536  
QY 150 -DEILLDLGHHERS-----IMTDLVYLSQTDGAG 177  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
537 LDAYMDFVFSKEDGGIFDNKNGLDVHLPVVGGTSKEPPLHIVHIAVEMAPIAKVGGIG 596  
QY 178 DWREKEAKDLTELVRITLYLONPKDCSK---ARKLVN----- 213  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
597 DVVTSLSRAVQL--NHNVDIVFPKYDCIKHNFVKDQFNRSYHMGGTGRIKWHGKVEGLS 655  
QY 214 -----INKGCGYGCQ-----LHHVWVCFMAYGTQRTLLILESONWRVATGAG 255  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
656 VYFLDPQNGLFQRCGVYGCADDAGFGFCFAALEFLQGGFHPD-ILHCHDMSAPVSW 714  
QY 256 ETVFRPVSECTDRSL--STGHSGEVNDKNIQVVELPIVDSLH----- 298  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
715 -----LFDKHYT-----QYGLIKTRIVFTIHNLEFGANAIGKAMTFAD 752  
QY 299 -----PRPYL-FLAVPELADRLRLRVHG-----DPAVW--WVSOPVKYLIRPQWLEKEI 346  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
753 KATTVSPYAKIKVAGNSVISAHLYKFHGIINGIDPDIMDPYNDNFI-----PVPYTSNV 807  
QY 347 -----EEATKLGFK-----HPVIGVHVRRTDKVGTEAAAFHPIEYV----- 383  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
808 VEGRAAKEELQNLRLGLKSADFPVVGITITLTHQKHLIKHAIWRTLRNGQVVLGSA 867  
QY 384 ---VHVEEHFQLIARMQV---DKKRVYLATDDPTLLKEAKTKYSNYEFISDNISWSAGL 438  
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
868 PDPRIQNDVFNLANQLHSHSGDRARLVLTIDEPL-----SHLIYAGADFILVPSIPEPCGL 923

RESULT 6  
T43527  
sp8 protein - fission yeast (Schizosaccharomyces pombe) (fragment)  
C:Species: Schizosaccharomyces pombe  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Feb-2000  
C:Accession: T43527  
R:Jimenez, M.; Goday, C.  
submitted to the EMBL Data Library, June 1997  
A:Description: Schizosaccharomyces pombe myosin-like.  
A:Reference number: 222545  
A:Accession: T43527  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1173 <JJM>  
A:Cross-references: EMBL:AF010473; PIDN:AAB65416.1  
A:Experimental source: strain 972h(-)  
C:Genetics:  
A:Gene: sp8

Query Match 3.8%; Score 118; DB 2; Length 1173;  
Best Local Similarity 19.9%; Pred. No. 8.7;

```

Qy      184   ---AKDITELVQRRTITYLQNPKDCSKARKLVNCINKGCGYCOLHHVVVCFMIAYGTOQT 240
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1326   LVSKENLIEELNQIHLKSELETVKSKSDLENER-----AQNSKI 1368

Qy      241   LILESQNRVATGWETVPFPVSETCTDRSGLSTGHSGEVNDKNIOVVLPIDVDSLHPR 300
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1369   EQLELKNTKLA-AAWRTKYBQVVNKSLEKINQIRQLQSORTSLEAKVAEHCQLNEQLNK 1427

Qy      301   PPVILPLAVPEDLRLLRVHGDPAAVMVSOFVKYLIRPQPWLKEIEEATKLGFKHPVI 360
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1428   PSATPTATQ-----SEPSTVSLBEFNS-----TKEELSSTORKULSEINDIL 1469

Qy      361   GVHVRRTDKV-----GTAAAFH-PIEYM-----VHVEEHFOLLARRMQVDKKRYVLA 407
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1470   NTTKEELEKVRQNSKNSEGTSKDTEIPNEEMERKKVMQOEVLRLRSIAKELQKNNELLR 1529

Qy      408   TDDPTLLKEAK----TKSYNYEFISONSISWSAGLNHRNTENSILRGVILDIHF----- 456
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1530   KQNVQLQDVKALQETVWSSEAEASVHADTKDLNLKKTTEMLSVTFQVIFNESISDF 1589

Qy      457   LSQADFLVCTFSQVCRAVEIMOTLHPDASANFH--SLDDI 496
          |||||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1590   STSTADF--TTFQKWEKREILQKDVEEQVAQSHQKQLDNI 1630

RESULT 8
F86161
F1003.10 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: F86161
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;
  Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; D'Ansen, N.F.; Hughes, B.; Huizar, L.
  Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.
  C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.;
  Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
  Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: F86161
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1744 <STO>
A:Cross-references: GB:AE0051172; NID:G4587570; PIDN:AAD25801.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match           3.8%; Score 117; DB 2; Length 1744;
Best Local Similarity 20.7%; Pred.No.18;
Matches 143; Conservative 98; Mismatches 225; Indels 224; Gaps 34;

Qy      35   HPD---HSSRELKILAKLERLKQONEDLRMA--ESLRPEGPIDQG----- 77
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     627   HPSPFGSSVKELOEENSKLKEITER-ESEKTLALIEKLEMWEKLQKNLLSENSIDLNA 685

Qy      78   ---TATGRVRLVEEQVLKAKEQ-----IENKYOARNGLGKHHEILR 116
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     686   ELETRIGKLTLEEASMSLAEKSGLSHKDKMLISRLQSATENSKK--LSEENVLE 740

Qy      117   RLRENGAKELWFFLOSELKKLK---HL-----EGNELQRHADEILLDGHERSI 163
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     741   NSLFANVEL--EELSKKLGSLEESCCHLLNDCKTTLTSRESLSLHDTM-----RKR 791

Qy      164   MTPLYLVSQTDGAGDWREKAEDL----TELVRRTIYLIQ-----NPDCSKARKL 210
          ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     792   IEDL-----EKHAELKVKVLELATERESSLOKIELGVSLNAKCEVASFV 838

Qy      211   VCNIKGCYGCYCOLHHV---YCFMIAYGTQ-----RTLTI 242
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||:

```



Db 839 QFSESRMNGMESTIHHLODENQCRVREYQVELDRAHDAHEIIVLQKCLQDWMLEKSSLI 898  
Qy 243 LESQNRVATGWTVPVSTCTDRSG-----LSTGHW----- 277  
Db 899 AENQDIKAS-----KLEKLVSELEENIGKQVQIDSSINCILRTGIYVLMKLEIIPG 955  
Qy 278 --SGEVNDKNIQVVELPIVDSLHPPYPLPLAVPEDLADRLRV-----HGDPVAVWVWSQ 330  
Db 956 IGSQDNGSRQDNMH-DILNRL-----EDMQTLLSIRDENQHSALLENLVIE 1002  
Qy 331 FVKYL-----LRQPMLEKEIEATKLGFK-----HPVIGVHVRRTDKVGTAAFHPI 379  
Db 1003 FLRQLKSEAVGIETSKKLEELSESCQCOQSFSDRDETQKLPVNGELTTKNGQVNRKV 1062  
Qy 380 EBYMVHVEE-HFOLLARR-----MQVDK-----KRVILATDDPTLLKEAKTKSYNEFIS 428  
Db 1063 --LWVEIEDPHRQVLOLRDDYTLQDNNKTLDKAYL-TKSTLQLEBEKCKLED----- 1114  
Qy 429 DNSISWSAGLHNRNTENSLRGVILD-----IHFLSQADFLVCTFSQVCRVAYEIMQ 480  
Db 1115 DISLLSETIYQNSLIILLEDVILEKSGAMKLNEDLRLSIVKCKLEEV-----RELGD 1170  
Qy 481 TLHPDASANFH-----SLDDIYFGGQNAHQIATVPHPKRTTEETPMPECDIIGVAG 533  
Db 1171 KLSADIANFQVVLKESNAELLARSANVHLEHIANVKVQKEKL-LEAMLMSIMQ 1229  
Qy 534 NHWDGYSKGNKGLKGTGLYPSYKVKREIE 563  
Db 1230 NEKSELSKAVE-----GLECKYKEAKALE 1253

RESULT 9  
D96796  
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96796  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;  
ansen, N.P.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D96796  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1871 <STO>  
A:Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F28016.15  
A:Map position: 1

Query Match 3.8%; Score 117; DB 2; Length 1871;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 45; Conservative 38; Mismatches 76; Indels 30; Gaps 8;  
Qy 24 YTGHLVRDNDHPDSSRLSK-ILAKLERLQKQEDLRMAESLRIPGPDQGTATCR 82  
Db 1065 HRTGG-----EDHNDHKEEQENVTAKAE-LNTEEDSFKAQVEIEKQHGELKRSVMQAK 1118  
Qy 83 VRLVEE---QVKAQEIQENYKQARNLGK-----DHEILRRRIENGAKELWF 128  
Db 1119 RQTEEKQKTRAMEKNETVERKQTKDGLKRGEDPELGGHE--RGGEEDRIEEL-- 1174  
Qy 129 FLQSEL-----KKLKHLEGNELQRADEILLDGLGHERSINTDLYLSQTDGAGWREKA 184  
Db 1175 -VETESIDHKEKVKKDEYILRSQDTGKVDLGERERRSKQRKIHKSVEIDGQDEDA 1233

Qy 185 KDLTELVR 193  
Db 1234 EEAATAVVS 1242

RESULT 10  
T30934  
myosin-like protein - Parascaris univalens  
C:Species: Parascaris univalens  
C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T30934  
R:Estebar, M.R.; Giovinarzo, G.; de la hera, A.; Goday, C.  
J. Cell Sci. 111, 723-735, 1998  
A>Title: PUMAL: a novel protein that associates with the centrosomes, spindle and centri  
A:Reference number: 220936; MUID:98139561; PMID:9472001  
A:Accession: T30934  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1955 <EST>  
A:Cross-references: EMBL:AF009623; NID:g3068589; PID:g3068590; PIDN:AAC38995.1  
C:Genetics:  
A:Note: pumal

Query Match 3.8%; Score 117; DB 2; Length 1955;  
Best Local Similarity 19.6%; Pred. No. 20;  
Matches 94; Conservative 80; Mismatches 181; Indels 124; Gaps 23;

Qy 41 RELSKILAKLERLQKQ-----NEDLRMAESLR--IPEGPIDQGTATGRVRLVLEQLVKA 93  
Db 788 RDLKDAMTEVEBELKKQIQKQDEENSERLESVLRTKISSDTVDTSS-----EIAEYTEVKV 841  
Qy 94 KEQIENYKQARNGLKGDKDHEILRRRIENGAKELWFFLQSELKCLKLH-----EGNEL-- 145  
Db 842 KELREKYKADLER-LQSNKQDLERRVQ-----ILLEDELAERQRIVERQRTENMDLKL 892  
Qy 146 --QRHADILLGLGHERSINTDLYLSQTDGAGWREKA---KDLTELVRRTIYQLON 200  
Db 893 EYQLESORLRAEMA-----TVELKYQSEVEDERQSRDADSWKVTSEELRSKISPFMEK 946  
Qy 201 PKDCSKARKLYCNKGGYQGLHHVYVCFMIAVGTORTLILES-----QN 247  
Db 947 MLEBAKHRETV--LREEATEWEEKHDI-----SNESLKLNEIRISDAEDIQK 996  
Qy 248 WRVATGGWETVIRPVSETC-TDRSGLSGTGH-----WSGEVNDKNIQVVELPIVDSLHPRP 301  
Db 997 WKDQVHAQNEIKNLERCETLRQLTAANDRVASLNTTINEQTSKIREL-----NSHEH- 1051  
Qy 302 PYLPLAVPEDLADRLRLRVHGDPAVWVVSQFVKYLIRPOPWLEKEIEBEATKKL--GFKHPV 359  
Db 1052 ----RLEEDLADSRATSSA-----IENDLGNATGRLSSEHNA 1086  
Qy 360 IGHVHRTDKVGTAAAFPIEYVHVVEHFQLLARRMQVDKRVLYLATDDPTLLKEAKT 419  
Db 1087 ILOSENRKSKTFEIALKHQIDTIM-NTKESCESEVERL---KKKIVQTT---TITKEQNE 1139  
Qy 420 KYSNVEFISDNISWSAGLHNRNTENSLRGVILDHFLSQADFLVCTFSQVCRVAYEI 478  
Db 1140 KIEKLRIEHD-----HLERYREKTEKVEDRLKEVE---KTFELKVNRAQEL 1183

RESULT 11  
B72765  
hypochemical protein APE0110 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 02-Feb-2001  
C:Accession: B72765  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;  
DNA Res. 6, 83-101, 1999  
A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: B72765  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-919 <KAW>  
A:Cross-references: DDBJ:AF000058; NID:g5103388; PIDN:BAA79020.1; PID:g5103499  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0110  
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 3.8%; Score 116.5; DB 2; Length 919;  
Best Local Similarity 22.5%; Pred. No. 7.8;  
Matches 108; Conservative 70; Mismatches 160; Indels 143; Gaps 25;

QY 50 LERLKQONEDLRMAESLIPGPIQGTATGRVRLVLEQLVKAKEIENYKQARNGLG 109  
DB 522 LSKNLQLEGBREL--GFQTP--DLAKAEQKLRLERLEAKLNSLEEKVRN--LS 575  
QY 110 KDEILR-----RRI---ENGAKELWFFLOSELKKLKHLGSELQRLORHADEILLDL 156  
DB 576 REEVALREAKTRALEVLQRLGKEBEAREKLTLSSESKLERM-----LVSKAEDLATRL 631  
QY 157 GHHERSIMTDLY-----YLSQTD---GAGDWREKEAKOL-----TELVRORITVL 198  
DB 632 GITAYRSLDDLEKAREALEGVDKLSAIERLEEARLKEEAALKWEAEQVMKLEEL 691  
QY 199 QNPKQCSKARKLVCMNKGCGYCOLHHVYVCFMAYGTQRTLI-LESQNWRYATGGWET 257  
DB 692 E-----AEKELRKEVSRKSETEARLKEV-----QNTLAELDDRISR----- 728  
QY 258 VFRPVSETCTDSGLSTGHSGEVNDKNVQVELPIVDSLHPRPPYLPPLA-----VPEDLA 313  
DB 729 IDRMGELQTRIREMKSRKASGE-----EALK-----YLPAAASRRIMEISG 771  
QY 314 D-----RLRVHGDPAVWVSQFVKYLIRPQPLEKEIEBATKLGFKHPVIGVHV-- 365  
DB 772 ETAYRLLAV-----LEBMDILSR--FNLDVAGVEIREKAA 807  
QY 366 ---RTDKVGTAEAFPIEYVMVHEHFQJ-----LARRMQVQKRVYLATDPTLLKEA 417  
DB 808 REIEVKAIGGAYRPLE--AVSGGERTVLAISFVLNKAVGGKLGFLALDEPTANLDE 865  
QY 418 KTKYSNYEFDISNSISWAGLHNRNTENSLRGVILDIHLSQADF--LVC--TFSSQVCR 473  
DB 866 DRRRLSIVELRGISV-----EGLVRQLVVVTHEDVRDYADTICLVTRTQQGSR 914  
QY 474 V 474  
DB 915 V 915

RESULT 12  
S02771  
myosin heavy chain A [similarity] - Caenorhabditis elegans  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Caenorhabditis elegans  
C>Date: 31-Dec-1993 #sequence\_revision 19-May-2000 #text\_change 19-Apr-2002  
C:Accession: T23622; S02771  
R:Harris, B.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19773  
A:Accession: T23622  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1992 <WIL>  
A:Cross-references: EMBL:Z78199; PIDN:CAB01576.1; GSPDB:GN00023; CESP:K12F2.1  
A:Experimental source: clone K12F2  
R:Diib, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.  
J. Mol. Biol. 205, 603-613, 1989  
A:Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gene  
A:Reference number: S02771; MUID:89178677; PMID:2926820  
A:Accession: S02771  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-116,140-1992 <DIB>

A:Cross-references: EMBL:X08067; NID:g6798; PIDN:CAA30856.1; PID:g6799  
C:Genetics:  
A:Gene: myo-3; CESP:K12F2.1  
A:Map position: 5  
A:Introns: 46/1; 192/1; 292/1; 468/2; 1921/3  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle coiled coil  
F:89-802/Domain: myosin motor domain homology <MMOT>  
F:202-209/Region: nucleotide-binding motif A (P-loop)  
F:690-712/Region: actin binding #status predicted  
F:793-807/Region: actin binding #status predicted  
F:875-1992/Domain: coiled coil #status predicted <COI>  
F:875-1189/Region: S2  
F:1190-1992/Region: light meromyosin  
F:153/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:208/Binding site: ATP (Lys) #status predicted  
F:730,740/Active site: Cys #status predicted

Query Match 3.8%; Score 116.5; DB 1; Length 1992;  
Best Local Similarity 19.1%; Pred. No. 23;  
Matches 89; Conservative 69; Mismatches 148; Indels 161; Gaps 19;

QY 42 ELSKILAKLERLKQONEDLRMAE-----SLRIPEGPIDOGTATGRVRV 85  
DB 954 KLSDI TQLEDMQERNEDLARQKKTDOELSDTKKHVDLESLRKAQEKQSRDHNR 1013  
QY 86 LEEQLVKAKEIENYKQARNGLGHDKHIEILRRRIENGAKELWFFLOSELKKLKHL--GN 143  
DB 1014 LODEMANQDEAVAKLKE-----KQHQ-----EESNRKLNEDLQSEEDKVNHEKIRN 1061  
QY 144 ELQRHADILLDLGHHERSIMTDLYLSQTDGADWRKEAKDLTELVRITVYLNQPKD 203  
DB 1062 KLEQOMDEL-----EENID-----REKRSRGDIEKAKRV-----EGD 1094  
QY 204 CSKARKLVCMNKGCGYCOLHHVYVCFMAYGTQRTLISSLQNWRYATGGWETVFRPV 263  
DB 1095 LKVAQENIDEITK-----QKHDV----- 1112  
QY 264 ETCTDRSGLSTGHSGEVNDKNVQVELP-IVDSLHPRPPYLPPLAVPBDL-ADRLLRVHG 321  
DB 1113 ETTLLKKEEDLHHTNAKLANNISIIAKLQRLIKELTARNAEI-----ESELEARNRQKS 1168  
QY 322 DPAVWVSQFVKYLIRPQPLEKEIEBATKLGFKHPVIGVHV-----R 365  
DB 1169 D-----RSRSEARELEELTELEQOGGATAQAQLEANKKREAIKLRREK 1214  
QY 366 RTDKVGTAEAFPIE---EYVMVHEEHFOLLAR-----RMQVQKRVYLATDPT 411  
DB 1215 EEDSLNHETALSSLRKRGDSVAELTEQLETLQKLKAKSEAKSKLQRLDLESQHTDSE 1274  
QY 412 T-----LLKEAKTKYSNYEFISD---NSISWSAGLHNRV-TENS 446  
DB 1275 VRSRQDLEKALKTIIEVQVSELQTKADEQSRQLQDFAAKLNRLNNNS 1321

RESULT 13  
H88187  
protein C18H9.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: H88187  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: H88187  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <STO>  
A:Cross-references: GB:chr\_II; PID:g722387; GSPDB:GN00020; CESP:C18H9.8  
C:Genetics:

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:39 ; Search time 10.5 Seconds  
(without alignments)  
2575.272 Million cell updates/sec

Title: US-09-971-773-24

Perfect score: 3085

Sequence: 1 MRAWTGSWRWMLILFWGT.....YKREKIEIVKYPTYPEAEK 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3074	99.6	575	1 FUT8_MOUSE	Q9wtc5 m alpha-(1,
2	3003	97.3	575	1 FUT8_HUMAN	Q9byc2 h alpha-(1,
3	2961	96.0	575	1 FUT8_BOVIN	Q9n0w2 b alpha-(1,
4	2910	94.3	575	1 FUT8_PIG	P79282 s alpha-(1,
5	126	4.1	997	1 SCF1_RAT	Q03410 rattus norv
6	125.5	4.1	845	1 SCF1_MESAU	Q60563 mesocricetu
7	120.5	3.9	751	1 COG5_DROME	Q9vj3d drosophila
8	119.5	3.9	1976	1 MYHA_BOVIN	Q27991 bos taurus
9	118.5	3.8	976	1 SCF1_HUMAN	Q15431 homo sapien
10	118	3.8	1727	1 ALM1_SCHPO	Q9utk5 schizosach
11	117	3.8	1955	1 FUM1_PARUN	Q61308 parascaris
12	116.5	3.8	919	1 RAS0_AERPE	Q9yifz1 aeropyrum p
13	116.5	3.8	1969	1 MYGA_CAEEL	P12844 caenorhabdi
14	115.5	3.7	1427	1 REST_HUMAN	P30622 homo sapien
15	114	3.7	214	1 OSF1_HUMAN	Q92882 homo sapien
16	113	3.7	879	1 RAS0_SULTO	Q95yrs sulfolobus
17	113	3.7	1057	1 CARB_STAAN	P58940 staphylococ
18	113	3.7	1217	1 ITN1_RAT	Q9wve9 rattus norv
19	113	3.7	3210	1 CENF_HUMAN	P49454 homo sapien
20	112.5	3.6	1270	1 ITN1_XENLA	Q42287 xenopus lae
21	112	3.6	1014	1 NEBL_HUMAN	O76041 homo sapien
22	111.5	3.6	1278	1 DYNA_HUMAN	Q14203 homo sapien
23	111.5	3.6	1281	1 DYNA_MOUSE	Q08788 mus musculu
24	111.5	3.6	1935	1 MYH9_HUMAN	P12883 homo sapien
25	111.5	3.6	1961	1 MYH9_RAT	Q62812 rattus norv
26	111	3.6	993	1 SCF1_MOUSE	Q62209 mus musculu
27	111	3.6	1597	1 CTR0_MOUSE	P49025 mus musculu
28	110.5	3.6	978	1 RAS0_AQUAE	Q67124 aquifex aeo
29	110	3.6	1057	1 CARB_STAAN	Q99ur5 staphylococ
30	110	3.6	1280	1 DYNA_RAT	P28023 rattus norv
31	109.5	3.5	1935	1 MYH7_RAT	P02564 rattus norv
32	109	3.5	352	1 SH32_MOUSE	Q62420 mus musculu
33	109	3.5	1230	1 UGS4_SOLTU	Q43846 solanum tub

## RESULT 1

FUT8\_MOUSE  
ID FUT8\_MOUSE STANDARD; PRT; 575 AA.

AC Q9WT52; Q921UL; 3.5 1934 1 MYH7\_MESAU

DT 28-FEB-2003 (Rel. 41, Created) 3.5 880 1 RAS0\_PIRAB

DT 28-FEB-2003 (Rel. 41, Last sequence update) 3.5 646 1 HS70\_THEAN

DT 15-SEP-2003 (Rel. 42, Last annotation update) 3.5 1976 1 MYHA\_HUMAN

DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-

DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)

DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alphas,6-fucosyltransferase)

DE (alpha1-6FucT) (Fucosyltransferase 8).

GN FUT8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_taxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=20358720; PubMed=10902914;

RA Hayaashi H., Yoneda A., Asada M., Ikeita M., Imamura T.;

RT "Molecular cloning of mouse alpha-1,6-fucosyltransferase and

RT expression of its mRNA in the developing cerebrum.";

RL DNA Seq. 11:91-96 (2000).

RN [2]

SEQUENCE FROM N.A.

MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to

CC the first GlcNAc residue, next to the peptide chains in N-glycans

CC (By similarity).

CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-

CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-

CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-

CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-

CC glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-

CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-

glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-N-acetyl-beta-D-glucosaminyl]asparagine.

-!- PATHWAY: Glycosylation.

-!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi (By similarity).

-!- SIMILARITY: Contains 1 SH3 domain.

-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AB025198; BAA76392.1; -;  
DR EMBL; BC010666; AAH10666.1; -;  
DR MGD; MGI:1858901; Fut8.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS0002; SH3; FALSE NEG.  
KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;  
KW Golgi stack; SH3 domain.  
FT DOMAIN 1 9  
FT TRANSMEM 10 30 CYTOPLASMIC (POTENTIAL).  
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT LUMENAL, CATALYTIC (POTENTIAL).  
FT DOMAIN 31 575  
FT DOMAIN 502 563 SH3.  
FT SITE 299 305 SH3-BINDING (POTENTIAL).  
FT SITE 365 366 IMPORTANT FOR DONOR-SUBSTRATE BINDING.  
FT CONFLICT 40 40 T -> S (IN REF. 2).  
FT CONFLICT 388 388 Q -> E (IN REF. 2).  
FT CONFLICT 418 418 N -> K (IN REF. 2).  
SQ SEQUENCE 575 AA; 66555 MW; 7BE2ED3146E0C45F CRC64;

Query Match 99.6%; Score 3074; DB 1; Length 575;  
Best Local Similarity 99.5%; Pred. No. 6.9e-203;  
Matches 572; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRAWTGSRWIMLILPAWGTLFYIGCHLVDRNDHPDHSRSLSKILAKLERLKQONEDL 60  
DB 1 MRAWTGSRWIMLILPAWGTLFYIGCHLVDRNDHPDHSRSLSKILAKLERLKQONEDL 60  
QY 61 RMAESLRIPGPIIDGTATGRVRLVEQLVKAKEIQENYKQARNGLGKDHILRRRIE 120  
DB 61 RMAESLRIPGPIIDGTATGRVRLVEQLVKAKEIQENYKQARNGLGKDHILRRRIE 120  
QY 121 NGAKELWFFLQSELKKLHLEGNELQORHADEILLDLGHHERSIMTDLYVLSOTDAGDWR 180  
DB 121 NGAKELWFFLQSELKKLHLEGNELQORHADEILLDLGHHERSIMTDLYVLSOTDAGDWR 180  
QY 181 EKEAKDLTVORRITLQNPDCSKARKLVNKNKGGCGQLHHVYVCFMAYGTQRT 240  
DB 181 EKEAKDLTVORRITLQNPDCSKARKLVNKNKGGCGQLHHVYVCFMAYGTQRT 240  
QY 241 LILESQNWRYATGGWETVFRPVSETCTDRSGLSGTHWSGEVNDKNIQVVELPIVDSLHPR 300  
DB 241 LILESQNWRYATGGWETVFRPVSETCTDRSGLSGTHWSGEVNDKNIQVVELPIVDSLHPR 300  
QY 301 PPVPLPAVEDLADRLRLVHGDPVWVWSQFVKYLRPOPVEKEIEEATKLGKHPVI 360  
DB 301 PPVPLPAVEDLADRLRLVHGDPVWVWSQFVKYLRPOPVEKEIEEATKLGKHPVI 360  
QY 361 GVHVRRTDKVGTAEAFPHPIEYVHVVEHFQLLARRMQVKKRVYLATDPTLLKEATK 420  
DB 361 GVHVRRTDKVGTAEAFPHPIEYVHVVEHFQLLARRMQVKKRVYLATDPTLLKEATK 420  
QY 421 YSNYEFISDINSISWSAGLNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCRAVEIMQ 480  
DB 421 YSNYEFISDINSISWSAGLNRYTENSRLRGVILDIHFLSQADFLVCTFSSQVCRAVEIMQ 480

QY 481 TLHPDASANFSLDDIYFYGQNAHNQIAVYPHKPRTEETPMPEPGDIIGVAGNHWGYS 540  
DB TLHPDASANFSLDDIYFYGQNAHNQIAVYPHKPRTEETPMPEPGDIIGVAGNHWGYS 540  
QY 541 KGINKRLGKTGLPSYKVKREKTIETVKYPTYPEAEK 575  
DB KGINKRLGKTGLPSYKVKREKTIETVKYPTYPEAEK 575

## RESULT 2

FUT8\_HUMAN  
ID FUT8\_HUMAN STANDARD; PRT; 575 AA.  
AC Q9BYC5; O00235; Q9BYC6; Q9P2U5; Q9P2U6;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein fucosyltransferase)  
DE (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
DE (alpha1-6FucT) (Fucosyltransferase 8).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.  
RX MEDLINE=97279058; PubMed=9133635;  
RA Yanagidani S., Uozumi N., Ihara Y., Miyoshi E., Yamaguchi N., Taniguchi N.;  
RT "Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 FucT) from human gastric cancer MKN45 cells.";  
RL J. Biochem. 121:626-632(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Embryo;  
RA Cailliau A., Balanzino L., Candellier J.J., Oriol R., Mollicone R.;  
RT "Differential splice variants of human FUT8 embryonic cDNA.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBAJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=20275614; PubMed=10814706;  
RA Yamaguchi Y., Ikeda Y., Takahashi T., Ihara H., Tanaka T., Sasho C., Uozumi N., Yanagidani S., Inoue S., Fujii J., Taniguchi N.;  
RT "Genomic structure and promoter analysis of the human alpha1,6-fucosyltransferase gene (FUT8).";  
RL Glycobiology 10:637-643(2000).  
RN [4]  
RP DONOR SUBSTRATE BINDING, AND MUTAGENESIS OF ARG-365 AND ARG-366.  
RX MEDLINE=20229550; PubMed=10764839;  
RA Takahashi T., Ikeda Y., Tateishi A., Yamaguchi Y., Ishikawa M., Taniguchi N.;  
RT "A sequence motif involved in the donor substrate binding by alpha1,6-fucosyltransferase: the role of the conserved arginine residues.";  
RL Glycobiology 10:503-510(2000).  
CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to the first GlcNAc residue, next to the peptide chains in N-glycans.  
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[N-acetyl-beta-D-glucosaminyl]asparagine = GDP + N(4)-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)]-[N-acetyl-beta-D-glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1->4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-N-acetyl-beta-D-glucosaminyl]asparagine.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:

```

CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9BYC5-1; Sequence=Displayed;
CC Name=2; Synonyms=Retinal;
CC IsoId=Q9BYC5-2; Sequence=VSP_001807, VSP_001808;
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D89289; BAA19764.1; -
CC EMBL; AF038281; AAB92372.2; -
CC EMBL; AF038280; AAB92372.2; JOINED.
CC EMBL; Y17979; CAA76988.1; -
CC EMBL; Y17976; CAA76985.1; -
CC EMBL; Y17977; CAA76986.1; -
CC EMBL; Y17978; CAA76987.1; -
CC EMBL; AB049828; BAB40975.1; -
CC EMBL; AB049740; BAB40929.2; -
CC EMBL; AB032573; BAA92859.2; -
CC EMBL; AB032572; BAA92858.1; -
CC EMBL; AB032568; BAA92858.1; JOINED.
CC EMBL; AB032569; BAA92858.1; JOINED.
CC EMBL; AB032570; BAA92858.1; JOINED.
CC EMBL; AB032571; BAA92858.1; JOINED.
CC PIR; JC5432; JC5432.
CC Genew; HGNC:4019; FUT8.
CC MIM; 602589; -
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; P55002; SH3; FALSE NEG.
KW Transfrase; Glycosyltransferase; Transmembrane; Signal-anchor;
KW Golgi stack; SH3 domain; Alternative splicing.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT LUMENAL, CATALYTIC (POTENTIAL).
FT DOMAIN 31 575
FT DOMAIN 502 563
FT SITE 299 305 SH3-BINDING (POTENTIAL).
FT SITE 365 366 IMPORTANT FOR DONOR SUBSTRATE BINDING.
FT VARSPLIC 280 329 EVKDKNQVVELPVIDSLHPRPPYLPVAPVEDLADRLVRVH
FT E (in isoform 2).
FT E (in isoform 2).
FT Missing (in isoform 2).
FT VARSPLIC 330 575 /FTID=VSP_001807.
FT MUTAGEN 365 365 R->A,K: COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 366 366 R->A,K: DECREASES ACTIVITY TO 3%.
SQ SEQUENCE 575 AA; 665315 MW; 5A24A93881E18D0 CRC64;
Query Match 97.3%; Score 3003; DB 1; Length 575;
Best Local Similarity 96.5%; Pred. No. 5e-198;
Matches 555; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
QY 1 MEAWTSGRWIMLILFANGTLLFYGGHLVRNDHPDHSRSLSKILAKLELKQONEDL 60
DB 1 MRPWTGSRWIMLILFANGTLLFYGGHLVRNDHPDHSRSLSKILAKLELKQONEDL 60
QY 61 RMAESLRPEGPIDQGTATGRVRLVEOLVKAKEQIENYKQAKNGLKDKHEILRRRIE 120
DB 61 RMAESLRPEGPIDQGTATGRVRLVEOLVKAKEQIENYKQAKNGLKDKHEILRRRIE 120
QY 121 NGAKELWFFLQSELKKLHLEGNLQRADEILDLGHHSRIMTDLVYLSQTDGAGDWR 180
DB 121 NGAKELWFFLQSELKKLHLEGNLQRADEILDLGHHSRIMTDLVYLSQTDGAGDWR 180

```

## RESULT 3

```

FUT8_BOVIN
ID FUT8_BOVIN STANDARD; PRT; 575 AA.
AC Q9N0W2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-
DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)
DE (GDP-L-Fucose:N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)
DE (alpha1-6FucT) (Fucosyltransferase 8).
GN FUT8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=20523969; PubMed=11070054;
RA Javard C., Dupuy F., Maftah A., Michalski J.-C., Oriol R.,
RA Petit J.-M., Julien R.;
RT "Ancestral exon organization of FUT8, the gene encoding the
RT alpha6-fucosyltransferase, reveals successive peptide domains which
RT suggest a particular three-dimensional core structure for the
RT alpha6-fucosyltransferase family.";
RL Mol. Biol. Evol. 17:1661-1672(2000).
CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to
CC the first GlcNAc residue, next to the peptide chains in N-glycans
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-{N-acetyl-beta-D-
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)}-[N-acetyl-beta-D-
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-
CC glucosaminyl]asparagine = GDP + N(4)-{N-acetyl-beta-D-
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)}-[N-acetyl-beta-D-
CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-
CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-
CC N-acetyl-beta-D-glucosaminyl]asparagine.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -!- TISSUE SPECIFICITY: Highest expression found in brain. Also found

```

CC in heart, lung, spleen and kidney.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF247186; AAF65460.1; -;  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00018; SH3; 1.  
 CC SMART; SM00326; SH3; 1.  
 CC PROSITE; PS50002; SH3; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 CC Golgi stack; SH3 domain.  
 CC DOMAIN 1 9  
 CC TRANSMEM 10 30  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC LUMENAL, CATALYTIC (POTENTIAL).  
 CC SH3.  
 CC SH3-BINDING (POTENTIAL).  
 CC SH3-IMPORT FOR DONOR SUBSTRATE BINDING (BY  
 CC SIMILARITY).  
 CC SEQUENCE 575 AA; 66492 MW; 604C66B8A5BF540 CRC64;  
 CC  
 CC Query Match 96.0%; Score 2961; DB 1; Length 575;  
 CC Best Local Similarity 95.0%; Pred. No. 3.7e-195;  
 CC Matches 546; Conservative 15; Mismatches 14; Indels 0; Gaps 0;  
 CC  
 CC QY 1 MRAWTGSGWIMILIFAWGTLLFYIGGHLVRDNDHSDHSSRELSKTLAKLELRKQONEDL 60  
 CC Db 1 MRPWTSWIMILIFAWGTLLFYIGGHLVRDNDHSDHSSRELSKTLAKLELRKQONEDL 60  
 CC QY 61 RMAESLRIPGEPIDQGTATGRVRLVEQLVKAKEIQENYKQARNGLGKDHHEILRRRIE 120  
 CC Db 61 RMAESLRIPGEPIDQGPASGRIRALEQLVKAKEIQENYKQTRNGLGKDHHEILRRRIE 120  
 CC QY 121 NGAKELWFFLOSELKXKLHLEGNELQSHADELLDGHHSRTIMTDLVYLSOTDGGDW 180  
 CC Db 121 NGAKELWFFLOSELKXKLHLEGNELQSHADEFLDGLHHSRTIMTDLVYLSOTDGGDW 180  
 CC QY 181 EKEAKDLTELQVRITYLQNPDKCSKARKLVNCKNGCGYQGLHVVVCFMIAVGTORT 240  
 CC Db 181 EKEAKDLTELQVRITYLQNPDKCSKARKLVNCKNGCGYQGLHVVVCFMIAVGTORT 240  
 CC QY 241 LILESQNRVYATGWTETFRPVSETCTDRSGLSTGHSGEVNDKNIQVVELPIVDSLHPR 300  
 CC Db 241 LILESQNRVYATGWTETFRPVSETCTDRSGVYTGHSGEIKDKNQVVELPIVDSLHPR 300  
 CC QY 301 PVLPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQWLEKEIEEATKLGKHPVI 360  
 CC Db 301 PVLPLAVPEDLADRLRVHGDPAVWVWSQFVKYLIRPQWLEKEIEEATKLGKHPVI 360  
 CC QY 361 GVHVRTDKVGTAAFPPIIEYVHVVEHFQILLARMQVDDKRVYLATDDPTLLKEAKTK 420  
 CC Db 361 GVHVRTDKVGTAAFPPIIEYVHVVEHFQILLARMQVDDKRVYLATDDPTLLKEAKTK 420  
 CC QY 421 YSNYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 CC Db 421 YPHYEFISDINSISWSAGLHNRNTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ 480  
 CC QY 481 TLHPDASANPHSLDDIYFGGQNAHQIAVYHPKPTBEEIPMEPGDIIGVAGNWDGYS 540  
 CC Db 481 TLHPDASANPHSLDDIYFGGQNAHQIAVYHPKPTADEIPMEPGDIIGVAGNWDGYS 540  
 CC QY 541 KGNRKLGTGLYPSYKVRKIEYKVPYVPEAK 575  
 CC Db 541 KGNRKLGTGLYPSYKVRKIEYKVPYVPEAK 575

RESULT 4  
 FUT8\_PIG STANDARD; PRT; 575 AA.  
 ID AC P79282;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-alpha-L-  
 DE fucosyltransferase) (GDP-fucose--glycoprotein fucosyltransferase)  
 DE (GDP-L-Fuc-N-acetyl-beta-D-glucosaminide alpha1,6-fucosyltransferase)  
 DE (alpha1-6FucT) (Fucosyltransferase 8).  
 GN FUT8.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 102-130; 333-344; 402-415 AND  
 RP 566-575, AND FUNCTION.  
 RC TISSUE=Brain;  
 RX MEDLINE=97066976; PubMed=8910378;  
 RA Uozumi N., Yanagidani S., Miyoshi E., Ihara Y., Sakuma T., Gao C.-X.,  
 RA Teshima T., Fujii S., Shiba T., Taniguchi N.;  
 RT "Purification and cDNA cloning of porcine brain GDP-L-Fuc:N-acetyl-  
 RT beta-D-glucosaminide alpha1--6fucosyltransferase.";  
 RL J. Biol. Chem. 271:27810-27817(1996).  
 CC -!- FUNCTION: Catalyzes the addition of fucose in alpha 1-6 linkage to  
 CC the first GlcNAc residue, next to the peptide chains in N-glycans.  
 CC The pH optimum is 7.  
 CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + N(4)-(N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-N-acetyl-beta-D-  
 CC glucosaminyl]asparagine = GDP + N(4)-(N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->3)-[N-acetyl-beta-D-  
 CC glucosaminyl-(1->2)-alpha-D-mannosyl-(1->6)]-beta-D-mannosyl-(1-  
 CC >4)-N-acetyl-beta-D-glucosaminyl-(1->4)-[alpha-L-fucosyl-(1->6)]-  
 CC N-acetyl-beta-D-glucosaminyl]asparagine.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
 CC form in trans cisternae of Golgi (By similarity).  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; D86723; BAAL3157.1; -;  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00018; SH3; 1.  
 CC SMART; SM00326; SH3; 1.  
 CC PROSITE; PS50002; SH3; 1.  
 CC Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;  
 CC Golgi stack; SH3 domain.  
 CC DOMAIN 1 9  
 CC TRANSMEM 10 30  
 CC CYTOPLASMIC (POTENTIAL).  
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 CC (POTENTIAL).  
 CC LUMENAL, CATALYTIC (POTENTIAL).  
 CC SH3.  
 CC SH3-BINDING (POTENTIAL).  
 CC SH3-IMPORTANT FOR DONOR SUBSTRATE BINDING (BY  
 CC SIMILARITY).  
 CC SEQUENCE 575 AA; 66229 MW; 0F199D0BC2018F7B CRC64;  
 CC  
 CC Query Match 94.3%; Score 2910; DB 1; Length 575;  
 CC Best Local Similarity 93.6%; Pred. No. 1.1e-191;



Matches 538; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MRATGSRWMLILFANGTLLFYGGHLVRDNDHSDHSSRELKSLAKLERLKOQNDL 60  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 1 MRPTGSRWMLILFANGTLLFYGGHLVRDNDHSDHSSRELKSLAKLERLKOQNDL 60  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 61 RMMAESLRIPGPIDOGTATGRVRLERQLVKAKEQIENYKQARNGLKGDKHEILRRRIE 120  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 61 RMMAESLRIPGPIDOGTATGRVRLERQLVKAKEQIENYKQARNGLKGDKHEILRRRIE 120  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 121 NGAKELWFFLOSELKXKLEHNEIQRHADEILDLGHHERSIMTDLYLVSOTDAGDWR 180  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 121 NGAKELWFFLOSELKXKLEHNEIQRHADEILDLGHHERSIMTDLYLVSOTDAGDWR 180  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 181 EKEAKDLTVORRITLYLQNPDKCSKARKLVNKNKGGYGCGLHHVYVCFMIAVGTORT 240  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 181 EKEAKDLTVORRITLYLQNPDKCSKARKLVNKNKGGYGCGLHHVYVCFMIAVGTORT 240  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 241 LILESQNRVYATGWTETFRPVSCTDRSGLSLTHWSGEVNDKNIQVVELPIVDSLHPR 300  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 241 LALESHNRYATGWTETFRPVSCTDRSGLSLTHWSGEVNDKNIQVVELPIVDSLHPR 300  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 301 PYPPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQPMLEKEIEBATKGLGPKHPVI 360  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 301 PYPPLAVPEDLADRLRVHGDPAVWVVSQFVKYLIRPQPMLEKEIEBATKGLGPKHPVI 360  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 361 GVHVRTDKVTEAAAFPIEEMVVEHFQILARMOVDKRVYLATDDPTLLKEATK 420  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 361 GVHVRTDKVTEAAAFPIEEMVVEHFQILARMOVDKRVYLATDDPTLLKEATK 420  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 421 YSNYEFISDNTSWSAGLHNRVTENSLRGVLDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 421 YPSYEFISDNTSWSAGLHNRVTENSLRGVLDIHFLSQADFLVCTFSQVCRVAYEIMQ 480  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 481 TLHPDASANFSLDDIYFGGQNAHQIAYVPHKPRTEEBEIPMEPGDIIGVAGNHDGYS 540  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 481 ALHPDASANFSLDDIYFGGQNAHQIAYVPHKPRTEEBEIPMEPGDIIGVAGNHDGYP 540  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 541 KGINKLGTGLYPSKYVREKLETYKTYTPEAK 575  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 541 KGWNKRLGTGLYPSKYVREKLETYKTYTPEADK 575  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 5  
 SCPI\_RAT  
 ID SCPI\_RAT STANDARD; PRT; 997 AA.  
 AC Q03410;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SYCP1 OR SCPI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=93099884; PubMed=1464329;  
 RA Meuwissen R.L.J., Offenberg H.H., Dietrich A.J., Riesewijk A.,  
 RA Iersel M., Heyting C.;  
 RT "A coiled-coil related protein specific for synapsed regions of  
 RT meiotic prophase chromosomes";  
 RL EMBO J. 11:5091-5100(1992).  
 CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE

SYNAPTONEMAL COMPLEX.  
 -!- TISSUE SPECIFICITY: Testis.  
 -!- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE  
 CELLS.  
 -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 DOMAIN HAS DNA-BINDING CAPACITY.  
 -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS  
 CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE  
 OTHER SPECIES SYCP1 SEQUENCES.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X67805; CAA4806.1; ALT FRAME.  
 Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 DNA-binding; Coiled coil.  
 DOMAIN 108 819 COILED COIL (POTENTIAL).  
 DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DOMAIN 982 990 ARG/LYS-RICH (BASIC).  
 SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;  
 -----

Query Match 4.1%; Score 126; DB 1; Length 997;  
 Best Local Similarity 20.6%; Pred. No. 0.65;  
 Matches 109; Conservative 78; Mismatches 180; Indels 162; Gaps 24;

Qy 42 ELSKILAKLEELKQONEDLRMAESLRIPGPIDOGTATGRVRLERQLVKAKEQIENY 100  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 155 ENEKSVLKLREIEQNKDL-----IKENNATRWCMNLLKETCARSAEKTSKY 201  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 101 KKQ-----ARNGLGK--DHEILRRRIENGAKELWFFLOSELKXKLEHNEIQR 147  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 202 EYERETQVYVLDNNIEKMLAFELRVQAEARLEMHFKLKHEDHEKIQHLE-EEYQK 260  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 148 HAD-----EILDLGHHERSIMTDLYL--SQTGDAGDWRKEBAKJUTELVQRRTYVQ 199  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 261 EVNNKENQVSLLIQSTEKENMKDLATFLLEESRDKANQLEK-----TKLDENLKELN 315  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 200 NPKDCSRKRLVNCINKCGYGCGLHHVYVCFMIAVGTORTLILLESQNRVYATGWTETV 259  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 316 EKDHLS-----ELEDIKMSMQRSMSQTKL---EEDLQIAT---KTIY 354  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 260 RPVSE-----TCTDRSGLSLTHWSGEVND----- 283  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 355 QLTEEKAQMEELNKAHTTHSLVTELKATCTLEELRTEQORLENEDQLKITMELQ 414  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 284 -----KNIQVVELPIVDSLHPRPYPPLAVPEDLADRLRVHGDPAVWVVSQ 330  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 415 KKSSELEBMTFKNKKEVELEELKTI-----LAEDQKLDELKQVEKLAELQKBEQ 466  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 331 FVKYLIRPQPMLEKEIEBATKGLGPKHPVIGVHVRPTDKVGTAAAFPIEEMVVEH 388  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 467 ELTFLLQTR---EKELHDL-----VQVTVTKTSE-----EHLVKQVEEMK 504  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 389 -----HFQLLARR---MOVDKRVYLATDDPTLLKEAKTKYSNVEFISDNTSWSA 436  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 505 TELESEKELKNIELTANSMLLENKLVQELASDMVLELKKHQEDINCKQOERMLKQIE 564  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 437 GLHNRVTENSLRGVLDI--HFLSQADFLVCTF--SSQVCR-VAYEIMQ 480  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Qy 565 TLEEK--EMNLRDELESVRKEFIQGDVEVKCKLDKSEENARSIEVLEK 611  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 6  
 SCPI\_MESAU  
 ID SCPI\_MESAU STANDARD; PRT; 845 AA.

AC Q60563;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome  
 DE synaptic protein) (Fragment).  
 GN SCPI OR SYNI.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 NCBI\_TaxID=10036;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Testis;  
 RC MEDLINE=95181577; PubMed=7876343;  
 RX Dobson M.J., Pearlman R.E., Karaïskakis A., Spyropoulos B.,  
 RA Moens P.B.;  
 RT "Synaptonemal complex proteins: occurrence, epitope mapping and  
 RT chromosome disjunction";  
 RL J. Cell Sci. 107:2749-2760 (1994).  
 CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE (BY SIMILARITY). HAS NON-  
 CC SPECIFIC DNA BINDING CAPABILITY.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX.  
 CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L32978; AAC42039.1; -;  
 DR PTR; 148176;  
 KW Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 KW DNA-binding; Coiled coil.  
 FT NON\_TER 1  
 FT DOMAIN <1 672 COILED COIL (POTENTIAL).  
 FT DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 753 756 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 830 838 ARG/LYS-RICH (BASIC).  
 SQ SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;  
 Query Match 4.1%; Score 125.5; DB 1; Length 845;  
 Best Local Similarity 19.7%; Pred. No. 0.57;  
 Matches 85; Conservative 81; Mismatches 161; Indels 105; Gaps 17;  
 QY 42 ELKILAKLERLKOQNEEDLRMAESRIPEGPID--QGTATGEVRVLEQLVKAKEQIEN 99  
 DB 288 ELKKILAEQKLDEKKQVEKLAEELQGEQELTLLQLTQREKVEHDLQSLVTKISDN 347  
 QY 100 YKQARNGLKGDHILRLRIENGAKELWFFLOSELKKLKH-----LEGNELOQRH 148  
 DB 348 YSKQV-----ELKTKLEE-----EKLKNAELTASCGLSLNNKLTQE 386  
 QY 149 ADEILDLGHERSIIMT----DIYLSQTDGADWREKAQDLTELIVQRRITVLPNQKDC 204  
 DB 387 TNDMALELKYQEDITNSKKQEBRLKQIENLEE-KETHLRDELESVRKEFIQOGNEVRC 445  
 QY 205 -----SKARKLVNINKGCGYCCQLHHVYCFMAYGTQRTLLILESQ--NRYATGGWE 256  
 DB 446 KLDKSEBARSIECEVLK-----KQKMKILEKCNKURKQA---E 483

QY 257 TVFRPVSETCTDRSGL--STCHWSEVNDKNIQVVELPI-VDSLHPR-----PPYLPLAV 308  
 DB 484 NKSYIEELHQENKALKSSAESKQLNAYEIKVKNKQLELESKAKQKQFQEMTDYQKIB 543  
 QY 309 PEDLADRLRLRVHGDPAVWVWSQFVKYLRPQPMLEKEIE-EATKKLGFKHPVIGVHVVRT 367  
 DB 544 VKKISEEKLGEVEKAKAMVDEAVK-----LQKEIDLRCQHKIAMVAMEKHKQY 595  
 QY 368 DKVGTG-----AAPHPTIEEYVMVVEEHFOLLARMQV---DKKRVYLAT 408  
 DB 596 DKVIEERDESLGCKNREQEQLSVKTALETLSNIRNELSVLKKQLETEREKEKLEK 655  
 QY 409 DDPTLLKEAKTK 420  
 DB 656 ENTALKDKDK 667  
 RESULT 7  
 COG5\_DROME  
 ID COG5\_DROME STANDARD; PRT; 751 AA.  
 AC Q9VJD3; Q8T0G3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Putative conserved oligomeric Golgi complex subunit 5 (Four way stop  
 DE protein).  
 GN FWS OR CG6549.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkely;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RX Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Voh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.":  
 RN Science 287:2185-2195(2000).  
 RL [2]  
 RP SEQUENCE OF 406-751 FROM N.A.  
 RX STRAIN=Berkeley; TISSUE=Embryo;  
 MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celnik S.E.;  
 RT "A Drosophila full-length cDNA resource";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -!- FUNCTION: Required for normal Golgi function (By similarity).  
 CC -!- SUBUNIT: Component of the conserved oligomeric Golgi complex which  
 CC is composed of eight different subunits and is required for normal  
 CC Golgi morphology and localization (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Golgi (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE COG FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF003655; RA053619.1; -;  
 DR EMBL; AY069335; AL39480.1; ALT\_INIT.  
 DR Flybase; FBgn0024689; fws.  
 KW Hypothetical protein; Transport; Protein transport; Golgi stack;  
 KW Membrane.  
 SQ SEQUENCE 751 AA; 84953 MW; E32B1D05BEEF6EBA CRC64;  
 -----  
 Query Match 3.9%; Score 120.5; DB 1; Length 751;  
 Best Local Similarity 20.7%; Pred. No. 1.1; Indels 177; Gaps 32;  
 Matches 127; Conservative 97; Mismatches 213;  
 QY 32 DNDPDHSS-----RELKSLAKLERLKQONEDLRMAESLRPEGP-IDQGTATGR 82  
 DB 19 DNDFTASMSHLITGOIQELSK-----QLQTKBELH---QOVRDKGALLQOATHAGR 69  
 QY 83 ----VRVLEBQLYAKE-----QIENYKQARNG---LGKDHILRRRIENGAKELWF 128  
 DB 70 FDAALNALAEDQVRVRETHRLKNQVDYQVQVENQTVLGRHDVSHLLRSAGT----- 124  
 QY 129 FLOSELKLLKHEGNELOHRADEILDIGH-HERSIMTDLVLSQTDGAGDWREKADL 187  
 DB 125 -LLSLTAKLKATK--DVLRLA-EIHPGLQIEDKELKDIDFIOQFRAYVVISSAQKIRNL 180  
 QY 188 TELVORRITYLO--NPKDCSKARKLCVNCIN-----KGCYGGCOL 224  
 DB 181 TQM--OLVTGLQERNENQVNAKLFPMNFTLEKSLDNLATFIADMEQSLKECFAGNDI 238  
 QY 225 H-----HWVCFMIAYGTQRT-LILESQNRVYATGGWET-----VFRPVSETCTDRSL 272  
 DB 239 SVLNKSPTHNVSKPAPSRGPKTQLTQTQNR--AKFWKSLHLLYDELFTCTQIKLL 296  
 QY 273 STG-----HWSGEVNDKNI-----QVVELPIVLSLHPRPPYPLVAVPEDLADRL 317  
 DB 297 KTALEQINQFGVTSSESDQCPQRFQVQVQLLRKSFDECPQHVTQTLOEGLSKLLTSAR 356  
 QY 318 ----RVHGDPAVWVSQFVKYLRQPWLEKEIBEATK-KLGFKHPVIGVHVRRYDKV-- 370  
 DB 357 GLEQLRHE-----PQFDELFPAP-----LEVGVVSKAANFKACLAGVDLPFGNETVDN 405  
 QY 371 ----GTEAFAHPIEYVMHV-----BEHFOLLAR-RMQVDKKEVYLATDDPTLL 414  
 DB 406 FIRVASTELSAALIDSLRLNATANVFAACGKELCTKLEAQIKLGADSKQV---VDLPNLQ 462  
 QY 415 KEAKTKYSNYEFISDMSISWSAGLHNRYTENSRLGVLDIHF-----LS 458  
 DB 463 QOQNTOLANLVF-----YYKDSVRRMLSDLHVHFEKTPGTAREIISRSLE 507

QY 459 QADFLVCTFSISQVCRAVE-----IMQTLHPDASANFHSLD-----DIYFGQNAH 505  
 DB 508 QADLLIGTIIQQIMESIITTSIIIVLSMHRPGLNSERMSTTGPSPMYMKELQEFVNRWS 567  
 QY 506 NQIAVYPHKVPTTEE 519  
 DB 568 HHIALFDQDK(MTKK 581

RESULT 8  
 MYHA\_BOVIN  
 ID MYHA\_BOVIN STANDARD; PRT; 1976 AA.  
 AC Q27991;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,  
 DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).  
 GN MYH10.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ohara M., Ishiguro N., Shinagawa M.;  
 RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 204-302 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=95301542; PubMed=7782316;  
 RA Itoh K., Adelstein R.S.;  
 RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle  
 RT myosin heavy chain II-B";  
 RL J. Biol. Chem. 270:14533-14540(1995).  
 CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
 CC CAPING (BY SIMILARITY).  
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB022023; BAA16494.1; -;  
 DR EMBL; U15716; AAA87715.1; -;  
 DR HSSP; P10587; 1BR2.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR001609; myosin head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF00603; myosin head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.  
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 FT Coiled coil; Alkylation; Multigene family.  
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.  
 FT DOMAIN 786 815 IQ.  
 FT DOMAIN 845 1976 COILED COIL (POTENTIAL).  
 FT NP BIND 178 185 ATP (POTENTIAL).  
 FT MOD RES 701 711 ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD RES 711 711 ALKYLATION (SH-2) (POTENTIAL).  
 SQ SEQUENCE 1976 AA; 22909 MW; 614354451COP790 CRC64;

Query Match 3.9%; Score 119.5; DB 1; Length 1976;  
 Best Local Similarity 25.7%; Pred. No. 4.5;  
 Matches 45; Conservative 33; Mismatches 44; Indels 53; Gaps 8;

QY 41 RELSKILAKLERLKQONEDLRMAESLRIPGPIQGTATGRVRLVEQLVKAQEQ---96  
 DB 1339 RQLEERSSLOEQOEEBEEARRSLEK-----LQALQAQLTDTKKKVVDDD 1383  
 QY 97 ---IENYKQARNGLGDHIELRRRIENGAKELWFFLQSELKHLKLEGNELQRAHDEIL 153  
 DB 1384 LGTIENL-EBAKKLKDVEVLSQRLKKA-----LAYDKLETK-----TRLQQLDLDLL 1433  
 QY 154 LDIGHERSIMTDL-----YYLSQTDGA-GDWREKAKDLT 189  
 DB 1434 VDL-DHQRTVSNLEKKQKQKFDQLABEKNISARYAEERDRAEAEREKETRLS 1487

## RESULT 9

SCPI\_HUMAN STANDARD; PRT; 976 AA.  
 ID SCPI\_HUMAN Q15431; O14963;  
 AC Q15431; O14963;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein 1 (SCP-1 protein).  
 GN SYCP1 OR SCPI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97294467; PubMed=911375;  
 RA Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J.,  
 RA Heyting C.;  
 RT "Human synaptonemal complex protein 1 (SCPI): isolation and  
 RT characterization of the cDNA and chromosomal localization of the  
 RT gene.";  
 RL Genomics 39:377-384(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=98037449; PubMed=9371399;  
 RA Kondoh N., Nishina Y., Tachida J., Koga M., Tanaka H., Uchida K.,  
 RA Inazawa J., Taketo M., Nozaki M., Nojima H., Matsumiya K., Namiki M.,  
 RA Okuyama A., Nishimune Y.;  
 RT "Assignment of synaptonemal complex protein 1 (SCPI) to human  
 RT chromosome ip13 by fluorescence in situ hybridization and its  
 RT expression in the testis.";  
 RL Cytogenet. Cell Genet. 78:103-104(1997).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC IS FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: Testis.  
 CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,

CC CC  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X95654; CAA64956.1; -;  
 CC EMBL; D67035; BAA22586.1; -;  
 CC Genbank; HGNC:11487; SYCP1.  
 CC MIM; 602162; -;  
 CC GO; GO:0005634; C:nucleus; TAS.  
 CC GO; GO:0005716; C:synaptonemal complex; TAS.  
 CC GO; GO:0003677; F:DNA binding activity; TAS.  
 CC GO; GO:0007131; P:meiotic recombination; TAS.  
 CC GO; GO:0007283; P:spermatogenesis; TAS.  
 CC GO; GO:0007129; P:synapsis; TAS.  
 CC DR Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 CC DNA-binding; Coiled coil.  
 CC KW DOMAIN 12 100 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 107 798 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 117 120 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT DOMAIN 679 682 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT DOMAIN 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT DOMAIN 961 969 ARG/LYS-RICH (BASIC).  
 CC FT CONFLICT 46 46 L -> F (IN REF. 2).  
 CC FT CONFLICT 106 106 F -> Y (IN REF. 2).  
 CC FT CONFLICT 153 153 F -> C (IN REF. 2).  
 CC FT CONFLICT 161 161 K -> T (IN REF. 2).  
 CC FT CONFLICT 168 168 E -> D (IN REF. 2).  
 CC FT CONFLICT 216 216 N -> S (IN REF. 2).  
 CC FT CONFLICT 225 226 HG -> FE (IN REF. 2).  
 CC FT CONFLICT 350 350 K -> N (IN REF. 2).  
 CC FT CONFLICT 360 360 E -> D (IN REF. 2).  
 CC FT CONFLICT 400 401 KN -> NY (IN REF. 2).  
 CC FT CONFLICT 406 406 K -> I (IN REF. 2).  
 CC FT CONFLICT 415 415 K -> T (IN REF. 2).  
 CC FT CONFLICT 449 449 E -> D (IN REF. 2).  
 CC FT CONFLICT 483 510 IOLTAITSQYKSEKVDKTELENEK -> YSYCHYHKV  
 CC FT CONFLICT 516 528 TVLPKRGQRKLSKRE (IN REF. 2).  
 CC FT CONFLICT 549 549 N -> I (IN REF. 2).  
 CC FT CONFLICT 560 560 K -> T (IN REF. 2).  
 CC FT CONFLICT 805 805 E -> D (IN REF. 2).  
 CC FT CONFLICT 941 941 P -> S (IN REF. 2).  
 CC SQ SEQUENCE 976 AA; 114069 MW; 8BA81D042AC2696B CRC64;

Query Match 3.8%; Score 118.5; DB 1; Length 976;  
 Best Local Similarity 19.5%; Pred. No. 2.1; Indels 139; Gaps 18;  
 Matches 91; Conservative 66; Mismatches 171;

QY 45 KILAKLELQKQONEDLRMAESLRIPGPIQGTATGRV-RVLBEQLVKAQEQENYKQ 103  
 DB 157 KVSLEKLEGIQENKDL-----IKENNAIRHLNLLKTCARSAEKTKKVE 203  
 QY 104 -----ARNGLKG---DHEILRRRIENGAKELWFFLQSELKHLKHLG-----NE 144  
 DB 204 REETQVYMDLNNNIEKMITAHGELRVAQENSRLMHPKLEKEDYEKIQHLEQVEKIND 263  
 QY 145 LQRADEILLDLGHHSIMTDLYVL---SQTDGADWRE-----K 182  
 DB 264 KEKQVSLLIQITSEKKN-MKDLTFLEESRDKVNQLEKTKLOSENKQIEKQHLLTK 322  
 QY 183 EAKOLTELIVQRRRI-TYLNQPKDCSKARKLVNINKGCGYGCQLHHVVYCFMAYGTQRTL 241  
 DB 323 ELEDIKVLSQRSVSTQKALEEDLQIATKTIQOLTE-----EKETQ 362  
 QY 242 ILESQNMRYATGGWETVPRVSVETCTDRSGLSLTHWSGEVNDKNIQVVELFIVSLHPRP 301

```

Db 363 MEESKARAAGFVVTEPE--TTVCSLELLRTEQORLEKNEKDQLKILTMELQKK----- 415
QY 302 PVLPLAVPEDLADRLRLRVHGDPAVWVVSQFVXLYLRPQPWLEKEIEEATKLGFKHPVIG 361
Db 416 -----SSELEEMTKLTNNKEVELEELKVLGKETTLLY 448
QY 362 VHVRRTRDKVGTAAFPHEEYVMVVEHPF--OLLARRMQVDKRVVL---ATDDPTLLKE 416
Db 449 ENKQ-----PEKTAELKGTQEOLIGLQAREKVEHDELTQLTATITSEQYYSKE 498
QY 417 AK-----TKYSNYEFISD--NSISWSAGLHNRVTENSRLRGVILDI 454
Db 499 VKDLKTELENEKLNKLTSLTSHCNKLS-----LENKELTQBTSDMTLEL 541

RESULT 10
ALM1 SCHPO
ID ALM1 SCHPO STANDARD; PRT: 1727 AA.
AC Q9UTK5; O13313; O9UTR8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1) _SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Guillermo R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Brown J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=20123449; PubMed=10660053;
RA Jimenez M., Petit T., Gancedo C., Goday C.;
RT "The alm1+ gene from Schizosaccharomyces pombe encodes a coiled-coil
RT protein that associates with the medial region during mitosis.";
RL Mol. Gen. Genet. 262:921-930(2000).
RN [3]
RP SEQUENCE OF 644-834 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,

```

```

RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
CC CYTOKINESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL133357; CAU62414.1; -
DR EMBL; AF010473; BAA65416.1; ALT_INIT.
DR EMBL; AB028012; BAA87316.1; -
DR PIR; T50073; T50073.
DR GeneDB SPombe; SPAC1486.04c; -
KW Coiled coil. 57 361 COILED COIL (POTENTIAL).
FT DOMAIN 443 463 COILED COIL (POTENTIAL).
FT DOMAIN 542 740 COILED COIL (POTENTIAL).
FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
SQ SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;

Query Match 3.8%; Score 118; DB 1; Length 1727;
Best Local Similarity 19.9%; Pred. No. 4.7; Indels 118; Gaps 19;
Matches 104; Conservative 78; Mismatches 223;

QY 34 DHPHSSRELSKILAKLERLKQONEDLRMAESLRIPGPDQGTATGRVRLVEQLVKA 93
Db 1166 NHVKRISKEMRIISQORQLLFLENKLRKTVESSNRVIADLRG-----ITEKDVST 1218
QY 94 KEQI---ENY-----KKQANGLGKDHKEI---LRRIENGAKELWFF----- 129
Db 1219 SESVGERSNYLNVALLNESKSLRENLERNEVEITELREKLETITKTLDFANRLKEQLE 1278
QY 130 --LQSELKKLHLEGN--ELORHADAELLDLGHHERSIWTDLYLSQTDGAGDWREKE-- 183
Db 1279 SOLQTEKAAVKKLENSNEYKRNHQLLSLN-----SSTSTSSDARLKNNE 1325
QY 184 ---AKDLTELVRRTYILQNPKDCSKARKLVCNINKGCGYCOLHHVVVYCFMAYGTORT 240
Db 1326 LVSKENLIEELNQELHGLKSELETVKSKSELENER-----AQNSKI 1368
QY 241 LILESQNWRYATGGWETVFRPVSETCTDRSGLSGTGHSNVDNKNIQVLELPIVDSLHPR 300
Db 1369 EQLEKLNKTLA-AAWRTKYEQVWVNSLEKHQIRQOLSQKTSLEAKVAECHQLNEQLNK 1427
QY 301 PPYPLAVPELDLADRLRLRVHGDPAVWVVSQFVXLYLRPQPWLEKEIEEATKLGFKHPVI 360
Db 1428 PSATPTATTQ-----SEPSTVSLBEFNS-----TKELSTQRLKSLIMDL 1469
QY 361 GVHVRRTRDKV-----GTEAAHPF--PTEEYV-----VHVEEHFOLLARRMQVDKRVVLA 407
Db 1470 NTKTEELKVRQNSNKSEGTSKDTEIFNEEMERKVMQOEVLRLRSIAKELQKNELLR 1529
QY 408 TDDPTLLKEAK-----TKYSNYEFISDINSISWSAGLHNRVTENSRLRGVILDIHF----- 456
Db 1530 KQNVQLQDVKALQETVVSSEAEASVHADTKDLENLKKTEEMLSVTFQVIFNESISDF 1589
QY 457 -LSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFH--SLDDI 496
Db 1590 STSTADF--TFVQKEWEKREIRILQKQVEEQVAQSHQQLDNI 1630

```



```

Db 522 LSLNQLBGLREL--GFQTPPE---DLAKAEQKRLMLRERLELRKLENSLEBKVRN-LS 575
Qy 110 KQHEILR-----RRI---ENGAKELWFFLOSELKKLKHLEGNELQORHADELILLD 156
Db 576 REVALREAKTRALEVQLRLGKKEEAREKLTLSSESKLERM-----LVSKAEADLATRL 631
Qy 157 GHHERSIMTDLV-----YLSQTD---GAGDWREKEAKOL-----TELVRORITVYL 198
Db 632 GITAYRSLDLDLEKAREALEGVVDKLSATERRLEEARLKEEAALKWEAEQVMKRLBEL 691
Qy 199 QNPCKDSKARKLVNINCKGCGVGCQLHHVYCFMAYGTCORTLI-LESQWRYAYGHWET 257
Db 692 E-----AEEKLRKEVSRSETEARLKEY-----ONTLAELEDDRISR----- 728
Qy 258 VFRPVSETCTDRSGSLGTGHSWGSEVNDKNIQVVELPIVDSLHPRPPVPLPLA-----VPEDILA 313
Db 729 IDREMGELQTRIREMKSRSKASE-----EALKL-----YLPAAASRRIMEIG 771
Qy 314 D-----RLLRVHGDPVAVMVVSQFVKYLIRPQPMLEKEIEEATKKLGPKHPVIGVHVR----- 365
Db 772 ETAYRELLAV-----LEDEMMDILSR--FNLDVAGVEIREKAA 807
Qy 366 ----RTDKVGTAEAFPIEYMHVHEHFOL-----LARRMQVDKKVYLATDDPTLLKEA 417
Db 808 REIEVKAIGGNGAYRPLE--AVSGGERTVLALSFLVALNKA VGGKLGFLALDEPTANLDE 865
Qy 418 KTKYSNVEFISDINSISWSAGLHNRYSLSIRGLVDIHLFSQADF--LVC--TFSSQVCR 473
Db 866 DRRRLSLVEIRGISV-----EGLVRQLVVVTHEDVRDYADTICLVTRTQOQSR 914
Qy 474 V 474
Db 915 V 915

RESULT 13
MYSAL CAEEL
ID MYSAL CAEEL STANDARD; PRT; 1969 AA.
AC P12824; Q21440.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain A (MHC A).
GN MYO-3 OR K12F2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
ON NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=89178677; PubMed=2926820;
RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
RT heavy chain gene family.";
RL J. Mol. Biol. 205:603-613(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Harris B.R.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE

```

```

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC C.ELEGANS
CC -!- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
CC WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X08067; CAA30356.1; -
CC EMBL; Z78199; CAB01576.1; -
CC PIR; T23622; S02771.
CC HSP; P08799; 1MND.
CC WormPep; K12F2.1; C312204.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR001609; myosin_head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS00096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 793
FT DOMAIN 794 923
FT DOMAIN 857 1969
FT NP_BIND 179 186
FT DOMAIN 667 589
FT DOMAIN 770 784
FT MOD_RES 130 130
FT MOD_RES 707 707
FT MOD_RES 717 717
FT CONFLICT 116 116
FT SEQUENCE 1969 AA; 225509 MW; 64577BBAF7EAD80A CRC64;
Query Match 3.8%; Score 116.5; DB 1; Length 1969;
Best Local Similarity 19.1%; Pred. No. 7.1;
Matches 89; Conservative 69; Mismatches 148; Indels 161; Gaps 19;
Qy 42 ELSKILAKLERLKQONEDLRMAE-----SLRPEGIDQQTATGRVTV 85
Db 931 KLSIDITGQLEDQMERNEDLARQKKTDOELSDTKKHQVDLESLRKAQEKQSRDHNIKS 990
Qy 86 LEEQLVKAKEQIENYKQARNGLGKHDEILRRRIENGAKELWFFLOSELKKLHLE--CN 143
Db 991 LQDENANQDEAVAKLNK-----KKHQ-----EESNRKLNEDLQSEEDKVNHLKIRN 1038
Qy 144 ELQRHADEILLDLGHHSRISMTDLYLSQTDGAGDWREKEAKDLTELVRITTYLQNPXD 203
Db 1039 KLEQOMDEL-----EENID-----REKSRGDIKAKRKV-----EGD 1071
Qy 204 CSKARKLVNINCKGCGVGCQLHHVYCFMAYGTCORTLI-LESQWRYAYGHWETFRPV 263
Db 1072 LKVAQENIDEITK-----QKHDV-----
Qy 264 ETCIDRSGLSTGHSWGSEVNDKNIQVVELP-IVDSLHPRPPVPLAVPEDL-ADRLLRVHG 321
Db 1090 ETLTKREEDLHHTNKLAEANSIIIAKLQRLKELTARNAL-----ELEEBAERNSRQS 1145
Qy 322 DPAVWVVSQFVKYLIRPQPMLEKEIEEATKKLGPKHPVIGVH-----R 365

```



Db 1146 D-----RSRSEARELEELTERLEQOGGATAAQLEANKREAIKLRREK 1191  
Qy 366 RTDKVGTAEAFHPIE-----EYVHVVEHFOLLAR-----RMQVDKRVYLATDDP 411  
Db 1192 EEDSLNHETAISSLRKHRSVAEUFEQUETLOKUKASEAKSKLORDLESQATDSE 1251  
Qy 412 T-----LLKEATKYNSYFISD---NSISWSAGLHNRV-TENS 446  
Db 1252 VRSQDLEKALKTIEVOYSELQTKADEQSRQLQDFAALKNRLNNS 1298  
RESULT 14  
REST HUMAN  
ID REST HUMAN STANDARD; PRT; 1427 AA.  
AC P30622;  
DC 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-  
Sternberg intermediate filament associated protein).  
GN RSN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Tissue=peripheral blood monocytes;  
RX MEDLINE=92289675; PubMed=1600942;  
RA Bilbe G., Delabie J., Bruegggen J., Richener H., Asselbergs F.A.M.,  
RA Carletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,  
RA de Wolf-Peters C., Shipman R.  
RT "Restin: a novel intermediate filament-associated protein highly  
expressed in the Reed-Sternberg cells of Hodgkin's disease."  
RL EMBO J. 11:2103-2113(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92405160; PubMed=1356075;  
RA PIERRE P., Scheel J., Rickard J.E., Kreis T.E.;  
RT "CLIP-170 links endocytic vesicles to microtubules."  
RL Cell 70:887-900(1992).  
CC -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN  
THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE  
CYTOSKELETON.  
CC -1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Names=Long;  
IsoId=P30622-1; Sequence=Displayed;  
Names=Short;  
IsoId=P30622-2; Sequence=VSP\_000765;  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS  
OF HODGKIN'S DISEASE.  
CC -1- SIMILARITY: Contains 2 CAP-Gly domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X64838; CAA46050.1; -  
DR EMBL; M97501; AAA35693.1; -  
DR PIR; S22695; S22695.  
DR Genew; HGNC:10461; RSN.  
DR MTM; 179838; -  
DR GO; GO:0005768; C:endosome; TAS.  
DR GO; GO:0005882; C:intermediate filament; TAS.  
DR GO; GO:0015630; C:microtubule cytoskeleton; TAS.  
DR GO; GO:0008017; F:microtubule binding activity; TAS.

DR GO; GO:0006899; P:non-selective vesicle transport; TAS.  
DR InterPro; IPR000938; CAP-Gly.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01302; CAP\_GLY\_2.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00845; CAP\_GLY\_1; 2.  
DR PROSITE; PS0245; CAP\_GLY\_2; 2.  
KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.  
FT DOMAIN 78 120  
CAP-GLY 1.  
FT DOMAIN 143 204  
SER-RICH.  
FT DOMAIN 232 274  
CAP-GLY 2.  
FT DOMAIN 304 331  
SER-RICH.  
FT DOMAIN 350 1342  
COILED COIL (POTENTIAL).  
FT DOMAIN 1408 1421  
CCHC-BOX.  
FT VARSPLIC 457 491  
Missing (in isoform Short).  
FT CONFLICT 1069 1069 D -> E (IN REF. 2).  
FT SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;  
Query Match 3.7%; Score 115.5; DB 1; Length 1427;  
Best Local Similarity 17.9%; Pred. No. 5.5;  
Matches 101; Conservative 97; Mismatches 208; Indels 157; Gaps 20;  
Qy 40 SRELSKILAKLERLKQONEDL-----RMAESLRIPGPIDQGTATGVR 84  
Db 588 SKENESLKLEHANKENSDVIALWKSLETAIAHQOAMEELKVFSFK-GLGTETAFA 646  
Qy 85 VLEQLVKAK-----EQIENYKKQARNGLGKHEILRRRIENGAKELWFFLOS----- 132  
Db 647 ELKQIEKWRLDYQHEIENLQNDQDSERAAHAKEMEARAKLMKVKEKENSLEAIRSKL 706  
Qy 133 -----ELKKLKHLEG--NELORHADEILLDLGHHSIMTDL 167  
Db 707 DKAEDQHLVEMEDTLNKLQEAIEIKVKELEVLQAKNEQTKVIDNFTSQLKATEKLL-DL 765  
Qy 168 YLSTQDAGDWREKAKDLTVELYRRTTYLQNPK--DCSKARKL-----VCNIN 215  
Db 766 DALRKASSEGSEKMKGRQQLAAEKQIKHLEIKNAESSKASSITRELOGLKLTNLQ 825  
Qy 216 KGCYGCQLHHVYCFMIAYGTQRTLIILESONRWYATGWTVPFVPSVETCTDRSGLSTG 275  
Db 826 ENLSEVSQVKETL-----EKELQILKEKFAEASEAVSVQSMQET----- 866  
Qy 276 HWSGEVDKNTQVVELPIVDSLHPPVPL-----AVPEDLADRLLRVHGDPVWVVS 329  
Db 867 -----VNKLHQEQFNMLSSDLEKLRNLADMEAKFREK-----D 902  
Qy 330 QVYKYLIRPQWLEKEIEATK-----KLGFKHPVIGVHVVRTDKVGTAA 375  
Db 903 EREEQLIKAKLEKLENDIAEIMKMSGDNSSQLTKONDELRLKRDVEELQLKLTKANENAS 962  
Qy 376 F--HPIEEYVHVVEHFOLLARRMVDKRVYLATDDPTLLKEAKTKSYNEFFIS---DN 430  
Db 963 FLOKSIEDMTVKAESQQAQKHEEKLEKRLSD--LEKKMETSHNQOQLKARYER 1020  
Qy 431 SISWSAGLHNRV-----TENSRL-----GVLDIHFL-SQADFLVCTSSQVC 472  
Db 1021 ATSETTKTHEEILQNLQKTLTDTEKLGAREENSGLLQLEELRKQAD---KAKAAQTA 1077  
Qy 473 RVAYEIMQTLHPDASANFHSLLDD 495  
Db 1078 EDAMQIMEQMTKEKTETLASUED 1100  
RESULT 15  
OSFL\_HUMAN  
ID OSFL\_HUMAN STANDARD; PRT; 214 AA.  
AC Q92882;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Osteoclast stimulating factor 1.  
GN OSTF1.





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 08:39:43 ; Search time 33.5 Seconds  
(without alignments)  
4429.261 Million cell updates/sec

Title: US-09-971-773-24  
Perfect score: 3085  
Sequence: 1 MRWTGSRWRMTLILFAWGT.....YKVRKTIETVKYPTYPEAEK 575

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organalle:\*  
9: sp\_plant:\*  
10: sp\_todent:\*  
11: sp\_virus:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2830.5	91.8	578	13 Q8AXS7	Q8axs7 xenopus lae
2	2187	70.9	446	4 Q8NEP2	Q8nep2 homo sapien
3	1480.5	48.0	619	5 Q9YVY5	Q9yvv5 d putative
4	978	31.7	559	5 O16882	O16882 caenorhabdi
5	810	26.3	169	4 Q8IUAS	Q8iuas homo sapien
6	670	21.7	512	5 Q81FW9	Q81fw9 ciona intes
7	601	19.5	560	5 Q81FW8	Q81fw8 ciona intes
8	130.5	4.2	1137	16 Q8F3E7	Q8f3e7 leptospira
9	129.5	4.2	1218	5 Q81ED2	Q81ed2 plasmodium
10	126	4.1	876	17 Q8TX14	Q8tx14 methanopyru
11	125	4.1	1025	10 Q8SAAS	Q8saas arabidopsis
12	122	4.0	745	11 P70565	P70565 rattus norv
13	121	3.9	462	16 Q8RIM2	Q8rim2 fusobacteri
14	120.5	3.9	3542	5 Q9U5M2	Q9u5m2 plasmodium
15	120	3.9	1177	16 Q8RCY8	Q8rcy8 thermoanaer
16	119.5	3.9	681	11 Q8BU18	Q8bu18 mus musculu

17	119.5	3.9	1200	11	Q921B9	Q921b9 mus musculu
18	118	3.8	2890	12	Q8B3M2	Q8b3m2 deformed wi
19	117	3.8	584	17	Q96YQ6	Q96yq6 sulfolobus
20	117	3.8	1744	10	Q9SA62	Q9sa62 arabidopsis
21	117	3.8	1871	10	Q9SRD5	Q9srd5 arabidopsis
22	116.5	3.8	795	4	Q9H2G7	Q9h2g7 homo sapien
23	116.5	3.8	804	4	Q8TC31	Q8tc31 homo sapien
24	116.5	3.8	975	13	Q98TQ5	Q98tq5 notothenia
25	116	3.8	674	5	Q18106	Q18106 caenorhabdi
26	116	3.8	1285	4	Q9UEG2	Q9ueg2 homo sapien
27	115.5	3.7	2376	5	Q9V5U0	Q9v5j0 drosophila
28	115.5	3.7	2376	5	Q966V1	Q966v1 drosophila
29	115	3.7	1137	5	Q93250	Q93250 caenorhabdi
30	115	3.7	1156	16	O66878	O66878 aquifex aeo
31	115	3.7	1397	10	Q9LP90	Q9lp90 arabidopsis
32	114.5	3.7	612	4	Q96JV2	Q96jv2 homo sapien
33	114.5	3.7	1014	5	Q8INU2	Q8inu2 drosophila
34	114	3.7	178	4	Q8WX78	Q8wx78 homo sapien
35	114	3.7	214	6	Q8MJ49	Q8mj49 sus scrofa
36	114	3.7	217	4	Q961J4	Q961j4 homo sapien
37	114	3.7	755	4	Q9UP84	Q9ufe4 homo sapien
38	114	3.7	1020	4	Q8MU84	Q8wu84 homo sapien
39	114	3.7	5362	2	Q93155	Q93155 bacillus su
40	113.5	3.7	1011	5	O61639	O61639 drosophila
41	113.5	3.7	1023	5	O76571	O76571 caenorhabdi
42	113.5	3.7	2003	5	O19658	O19658 caenorhabdi
43	113.5	3.7	2003	5	O22869	O22869 caenorhabdi
44	113	3.7	717	4	Q96ED9	Q96ed9 homo sapien
45	113	3.7	1094	5	O61618	O61618 drosophila

## ALIGNMENTS

## RESULT 1

Q8AXS7 PRELIMINARY; PRT; 578 AA.  
ID Q8AXS7  
AC Q8AXS7; 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative alpha-6-fucosyltransferase (EC 2.4.1.68).  
GN FUT8.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mollicone R., Michalski J.C., Bauvy C., Cailleau-Thomas A., Oriol R.;  
RA Candelier J.J., Martinez-Duncker I., Breton C., Codogno P., Oriol R.;  
RT "Splice variants of alpha6-fucosyltransferase are expressed early in human embryogenesis."  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ514872; CAD55853.1; -  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 578 AA; 66578 MW; 2741A03BDB3C5265 CRC64;

Query Match 91.8%; Score 2830.5; DB 13; Length 578;  
Best Local Similarity 89.3%; Pred. No. 3.6e-203;  
Matches 516; Conservative 36; Mismatches 23; Indels 3; Gaps 1;  
QY 1 MRWTGSRWRMTLILFAWGTLLFYGGHLVRDNDHPDHSRSLSKILAKRLKQNNEDL 60  
Db 1 MRWTGSRWRMTLILFAWGTLLFYGGHLVRDNDHPDHSRSLSKILAKRLKQNNEDL 60  
QY 61 RMAESLRIPGPIIDQGTATGVRVLEPOLVKAKEQIENYKQARN---GLGKHEILRR 117  
Db 61 RMAESLRIPGPIIEQQAAGRIRALEQLLKAKEQIENYKQSSNAVSGJGKHEILRR 120  
QY 118 RIENGAELWFFLOSELKKLHLENEIQHAEDEILLDLGHHSRISIMTDLYLSOTGAG 177

```
Db 121 ATEANGAFWYLVQSEVKKLKHLDNELQHVDEILIDMGHQORSVMTDLYYLSDTDGAG 180
QY 178 DMREREAKDLTELVRRTYLYLQNPDKCSKARKLVNKNKGGYGCQLHHVVCYCFMAYGT 237
Db 181 DMREREAKDLTELVRRTYLYLQNPDKCSKARKLVNKNKGGYGCQLHHVVCYCFMAYGT 240
QY 238 QRTLLLESQWRVATGNETVPRVSECTCTDRSGLSTGHWSGVNDKNTQVVELPIVDSL 297
Db 241 QRTLLLESQWRVATGNETVPRVSECTCTDRSGLSTGHWSGVNDKNTQVVELPIVDSL 300
QY 298 HRPBPPLPLAVPDLADRLRLRHGDPVAVWVSQFVKYLRPOPWLEKEIEEATKGLGPKH 357
Db 301 HRPBPPLPLAVPDLADRLRLRHGDPVAVWVSQFVKYLRPOPWLEKEIEEATKGLGPKH 360
QY 358 PVIQVHVRRTDKVGTAAFPHEEYVMHVEEHFQILLARMQVDKRVYLATDPTLLKEA 417
Db 361 PVIQVHVRRTDKVGTAAFPHEEYVMHVEEHFQILLARMQVDKRVYLATDPTLLKEA 420
QY 418 KTKYSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYE 477
Db 421 KAKYPOYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYE 480
QY 478 IMQTLHPDASAFHSLDDIYFVGQNAHNOIAVYHPKPTREEIPMEPGDIIIGVAGNHW 537
Db 481 IMQTLHPDASAFHSLDDIYFVGQNAHNOIAVYHPKPTREEIPMEPGDIIIGVAGNHW 540
QY 538 GYSKGINRKLGTGLTGLPSYKVKREKTIETVYKPYPEAEK 575
Db 541 GYSKGINRKLGTGLTGLPSYKVKREKTIETVYKPYPEAEK 578

RESULT 2
Q8NEP2
ID Q8NEP2 PRELIMINARY; PRT; 446 AA.
AC Q8NEP2;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE Putative alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-
DE alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUCT6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Buser D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dougan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jaitani D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
```

```
QY 401 KRVYLATDDPTLLKEATKYSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQ 460
Db 272 KRVYLATDDPTLLKEATKYSNYEFISNSISWSAGLHNRVTENSLRGVILDIHFLSQ 331
QY 461 DFLVCTSSQVCRVAYEIMQTLHPDASAFHSLDDIYFVGQNAHNOIAVYHPKPTREE 520
Db 332 DFLVCTSSQVCRVAYEIMQTLHPDASAFHSLDDIYFVGQNAHNOIAVYHPKPTREE 391
QY 521 IPMEPGDIIIGVAGNHWGYSKGINRKLGTGLTGLPSYKVKREKTIETVYKPYPEAEK 575
Db 392 IPMEPGDIIIGVAGNHWGYSKGINRKLGTGLTGLPSYKVKREKTIETVYKPYPEAEK 446

RESULT 3
Q9VVV5
ID Q9VVV5 PRELIMINARY; PRT; 619 AA.
AC Q9VVV5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative alpha-(1,6)-fucosyltransferase (EC 2.4.1.68) (Glycoprotein 6-
DE alpha-L-fucosyltransferase) (GDP-fucose-glycoprotein
DE fucosyltransferase) (GDP-L-Fuc:N-acetyl-beta-D-glucosaminide alpha1,6-
DE fucosyltransferase) (Alpha1-6FucT).
GN FUCT6 OR CG2448.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Buser D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dougan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jaitani D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
```







RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL944509; CAD52328.1; -  
SQ SEQUENCE 1218 AA; 143059 MW; EIDBD78388BC7ESB CRC64;  
  
Query Match 4.2%; Score 129.5; DB 5; Length 1218;  
Best Local Similarity 18.4%; Pred. No. 1;  
Matches 87; Conservative 88; Mismatches 157; Indels 141; Gaps 18;  
  
QY 32 DNDPHSSRELKILAKLERLQKQEDLRMAESLRIPGPDQ-----GTATG 81  
DB 707 DNE-----KLKESEKLSLEKAEKKIISIKELQIYENNLINMETSKYGVNR 760  
  
QY 82 RVRVLEOLVKAEOI-ENYKQAR-----NGLGKD---HEILRRRIENGAKELWFFLOS 132  
DB 761 KIEHKNEDKGNELSELYKEOKKUTVIRKLEKDISYEANKOKKEEDLKEIKKLON 820  
  
QY 133 ELKKLHLEGNELQRLHADEILLDLGHHSIMTDLVYLSQTDGAGDWREKEAKDLTE--- 189  
DB 821 KIQLE-TEHKKKEEIDVLLQIENYKQKKEETNLSSTDEIINEIEKKTIEDIEKNIN 879  
  
QY 190 LVORRTYLPONKDCSKARKLVNINKGCGQLHHVYVCFMIAGTORTLILESQNR 249  
DB 880 ITKENLKELEN-----KITELQSFSS---SYENEMKHVVKKI----- 913  
  
QY 250 YATGGWETVPRPVSETCTDRSGLSTGHSGEVNDKNIQVVELPIVDSLHPPPLPLAVP 309  
DB 914 -----EDLEKKSENLID-----LKKLENTLLD-----LQ 938  
  
QY 310 EDLADRLLRVHGDPVWVWSQFVKYLIRPOPWLEKEIEEATKKL-----GFKHPVIGVH 363  
DB 939 KDLKTS-----SDTVKLYLKHVHWIESVEPLFNKKYTPYDFENFRHVIQKK 985  
  
QY 364 VRTDKVGTAAAPHPIEYVWVVEHPQ-----LLARMQV---DKKRV----- 404  
DB 986 IQALONEQNKLIN-INKKAVQMEYQVQVDYKDLTKKSQVEBKKKIQEVIALDLVKKS 1044  
  
QY 405 -----YLATDDPTLLKEAKTKSYNYEFISDINSISWSAGLHNRYTEN 445  
DB 1045 ESSLAMVQINNEYFOAIFSTLLNNAQKLSIVDGLDIANGIEMKIAPNNWNKES 1097

## RESULT 10

Q8TXI4  
ID Q8TXI4 PRELIMINARY; PRT; 876 AA.  
AC Q8TXI4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE SWC1-family ATPase involved in DNA repair.  
GN SBCC OR MK0690.  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
OC Methanopyrus.  
OX NCBI\_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014;  
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natalie D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozayavkin S.A.;  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
RT and monophyly of archaeal methanogens."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
DR EMBL: AE010362; AAC01904.1; -  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_Transporter.  
DR InterPro; IPR003405; SMC\_C.  
DR InterPro; IPR003395; SMC\_N.

## RESULT 11

Q9SAA5  
ID Q9SAA5 PRELIMINARY; PRT; 1025 AA.  
AC Q9SAA5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE P25C20.13 protein.  
GN P25C20.13  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosida II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Lee J.M., Kremenetskaia I., Luros J., Ngan I., Liu A.,  
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,  
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,  
RA Walker M., Davis R.W., Ecker J.R., Federapfel N.A., Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC P25C20 sequence."  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR InterPro; IPR003395; SMC\_N.









**THIS PAGE BLANK (USPTO)**